UNCLASSIFIED

AD 651 792

PSYCHOLOGICAL RESEARCH ON OPERATIONAL TRAINING IN THE CONTINENTAL AIR FORCES

Meredith P. Crawford, et al

Army Air Forces
Washington, D.C.

1947

UNCLASSIFIED

AD-651792

Psychological Research on Operational Training in the Continental Air Forces

REPORT NO. 16

Army Air Forces
Aviation Psychology Program
Research Reports

Psychological Research on Operational Training in the Continental Air Forces

REPORT NO. 16

Edited by
MEREDITH P. CRAWFORD
Professor of Psychology and Dean of the Junior College
Vanderbilt University
RICHARD T. SOLLENBERGER
Associate Professor of Psychology
Mt. Holyoke College
LEWIS B. WARD
Research Associate
Harvard University
CLARENCE W. BROWN
Associate Professor of Psychology
University of California
EDWIN E. GHISELLI
Associate Professor of Psychology
University of California

1947

Preface

This report is a summary of research accomplished on operational training by aviation psychologists in the Continental Air Forces during the latter part of World War II.

In the presentation of this report there are two related objectives. First, an attempt will be made to present a record of the research activities of aviation psychologists in the Continental Air Forces; second, it is hoped to provide a summary of psychological research on operational training that will be of value to future research workers in this field. Additional and basic information relative to this report may be found for the most part in the files of the Psychological Branch, Office of the Surgeon, Headquarters, Strategic Air Command, Andrews Field, Camp Springs, Maryland.

Credit for the data included in this report is due primarily to the directors and staffs of the separate research projects in each of the continental air forces. These projects were established and functioned directly under Headquarters, Army Air Forces for several months prior to the activation of Headquarters, Continental Air Forces. The data presented are essentially consolidations of data provided by research personnel of each of several air forces. In the report no attempt has been made to assign special credit to individuals since, for the most part, all participated as a team in the activities carried on. A complete roster of all psychological research personnel involved appears in Chapter 1, table 1.1 of this report.

The planning and execution of the final report has also been a cooperative effort. The outline and framework of the report were developed at a conference of representatives of research personnel of all four air forces and of Headquarters, Continental Air Forces. The original writing of most of the chapters was delegated to the staffs of the various air forces. Final revision and editing was accomplished by the research personnel of the Psychological Branch, Office of the Surgeon, Headquarters, Continental Air Forces. Appreciation is also due personnel of the Psychological Branch, Office of the Air Surgeon, Headquarters, Army Air Forces, who read all of the chapters of this report and made many helpful suggestions and criticisms.

LEWIS B. WARD,
Lt. Col., A.C.

Bolling Field, D.C., 1 June 1946

Contents

CHAPTER ONE

Introduction

This summary of psychological research in the Continental Air Forces was prepared as one of a series of reports on the activities and findings of the Aviation Psychology Program in the Army Air Forces. While the material presented herein was gathered entirely within the Continental Air Forces, its significance may best be appreciated in relation to the research of aviation psychologists in other commands, which is treated in detail in other reports in this series. Since psychologists were assigned to the Continental Air Forces relatively late in the war, their mission was largely determined by the work already done in the Aviation Psychology Program during the previous 3 years.

In this introductory chapter there is first presented, for readers unfamiliar with operational training, a brief statement of the history, organization, and training function of the Continental Air Forces. This is followed by a statement of mission, organization, personnel and operating procedures of the psychologists within the command. Finally, there is a discussion of certain characteristics of training in the Continental Air Forces which offered special difficulties to aviation psychologists in the fulfillment of their mission.

THE CONTINENTAL AIR FORCES

History and Organization

The origin of the Continental Air Forces may be traced to the reorganization, in 1935, of the Army units charged with training of aviation personnel and employment of aircraft. In the reallocation of missions, the General Headquarters, Air Forces was established with the responsibility for training combat organizations in aerial warfare. The training of individuals in various specialties, both aerial and ground, remained under the direction of the Office, Chief of the Air Corps. Later, the AAF Training Command was established and took over the mission of individual training. The First, Second, Third and Fourth Air Forces, which make up the bulk of the present Continental Air Forces, grew out of the commands under General Headquarters, Air Forces. These four air forces were activated in March 1941, with the primary mission of defense of the continental United States. With the declaration of war on 8 December 1941, the First and Fourth Air Forces became almost exclusively engaged in the defense of the eastern and western parts of the United States respectively, while the Second and Third were assigned the training of combat units. As the need

for air defense of the continent became less important, all four air forces became engaged in training.

During the last 2 years of the war, the four Continental Air Forces occupied approximately the following sections of the country. The First Air Force extended along the eastern seaboard from Massachusetts to Georgia with headquarters at Mitchel Field, Hempstead, Long Island. The Second Air Force, with headquarters at Colorado Springs, Colorado, covered the Plains and Rocky Mountain states together with western Texas, New Mexico and Arizona. The Third Air Force extended throughout the Southeast and was administered from Tampa, Fla. The Fourth Air Force had bases along the west coast with headquarters at San Francisco, California. Each of these air forces had certain subordinate wings which specialized in training on different types of aircraft. During the last half of 1944 and the first half of 1945, training on different types of airplanes was distributed as follows. The First Air Force was responsible for medium (B–25 and B–26) and heavy (B–24) bombardment training, and for fighter training using P–47 and P–47N (very long range) airplanes. The Second Air Force carried on P–40 and P–47 fighter training, heavy bombardment training using both B–17's and B–24's, and very heavy bombardment training employing B–29's. The Third Air Force included wings responsible for fighter training using P–51's, heavy bombardment training using B–17's and during 1945, very heavy bombardment training with B–29's. The Fourth Air Force gave training in P–38 fighter aircraft, P–61 night-fighter aircraft, and heavy bombardment using B–24's.

By the end of 1944 all four air forces and the First Troop Carrier Command were primarily engaged in training for overseas duty. This similarity of functions resulted in the establishment of a single administrative and operating headquarters for the five commands. On 15 December 1944, Headquarters, Continental Air Forces was activated, to be opened on 1 April 1945 in its temporary headquarters at Bolling Field, D. C. The mission of this headquarters was to operate three general types of activities: training for overseas duty of new personnel received from the AAF Training Command, redeployment of personnel and units from Europe to the Pacific, and air defense of the continental United States. Policy making for the commands under this headquarters remained at Headquarters, Army Air Forces.

General Training Procedures

Operational training in the Continental Air Forces typically began with the receipt of individuals, trained in their aerial or ground specialties, from schools within the AAF Training Command. In operational training, these individuals were fitted into combat organizations, taught to use combat-type equipment and instructed in the basic tactics of aerial warfare. Upon completion of this training, they were shipped to the combat zone where they either began combat operations immediately or received further train-

ing in the particular tactical doctrines employed in the theater to which they were assigned. In general, training in the Continental Air Forces was designed to simulate as closely as possible the conditions of aerial warfare.

The basic requirements to be completed in operational training were stated in publications called Training Standards, originally prepared at Headquarters, Army Air Forces. After its establishment, Continental Air Forces carried on this function. The Training Standards set forth the minimum requirements to be accomplished in each type of operational training. These were usually stated in terms of number of hours to be spent in flying and ground training, and listed the subjects, types of maneuvers and aerial exercises to be covered. The headquarters of the four air forces amplified these training standards in appropriate training manuals which constituted syllabi for the courses included in the various types of training. The training requirements and procedures were further amplified by subordinate training wings, and particular schedules for each class were drawn up at the station level. In general, training schedules prepared at lower echelons included somewhat more work than the minimum specified in training standards from Headquarters, Army Air Forces. This extra work was completed as time, weather and maintenance conditions permitted.

Operational training was accomplished according to two general plans. The older plan, referred to as OTU (Operational Training Unit) training, was used to get an entire aerial combat organization ready for overseas shipment. This included the training of both air and ground echelons at all levels of command from the bomber command to the squadron. On an OTU base, trainee personnel included the complete complement of a group made up of three squadrons. The ground and flying training of this personnel was the responsibility of the Director of Training, a member of the permanent party of the base, assisted by permanent party instructors. Flight and squadron commanders and other supervisory personnel of the group assisted in this training, gradually taking over the responsibility toward the end of the period. Approximately four months were required to accomplish minimum training. While subordinate groups were being trained at various stations, their higher echelon command organization, wing and bomber or fighter command, was based at a central location, and was engaged in perfecting its organization and working out tactical exercises.

After overseas air forces were in operation, training was largely done in CCTS (Combat Crew Training Station) in order to provide replacement crews to keep the overseas organizations up to strength. The same training directives were used as in OTU training, but trainee personnel comprised only combat crews. The Director of Training and instructors were permanent party personnel and typically remained at the station through successive classes. Normally, at any given time there were present on a base either two or three classes in different stages of training. The CCTS training program was more typical, in terms of organization plan, of that in the AAF Training Command than was OTU Training, since successive classes

were instructed by the same personnel, and a more stable training staff could be maintained. When combat crews completed this training, they were sent through a staging area to the overseas units needing replacements.

Toward the close of the war in Europe, a third type of training was undertaken by the air forces in connection with the redeployment program. Combat organizations were returned intact from the European theater, replacements were added if needed, and the complete organization was retrained for duty in the Pacific. The general plan of this training followed that of the OTU training described above. By the time redeployment got under way, psychological research sections were occupied with other studies, so that no personnel were available to obtain data in this program.

ASSIGNMENT OF RESEARCH PERSONNEL

Mission

General Mission

Aviation psychologists were assigned to the Continental Air Forces for the primary purposes of making follow-up studies of aircrew members undergoing operational training. At this time, the psychological program had already been in operation for 2½ years in the AAF Training Command, where a battery of psychological tests was administered to all aircrew candidates for selection and for classification in aircrew specialties. The results of these tests had been extensively validated against success or elimination from Training Command Schools. To follow these individuals through operational training represented the next logical step in validating the selection and classification procedures. Because Continental Air Forces training simulated combat conditions, and at the same time offered data which were more accessible than any which could be obtained in the theaters, follow-up studies in the training air forces were regarded as essential. Further, it was believed that psychologists could furnish useful information on aircrew personnel by making available psychological records from the AAF Training Command and by devising and administering new tests for special types of assignments. On 23 August 1944 a letter was addressed to the Commanding Generals of the First, Second, Third and Fourth Air Forces, subject "Research Officers," which assigned psychological personnel to these commands and included the following general statement of mission:

a. To collect, assemble, and make available to commanding officers, classification test scores, stanines, proficiency test scores, training records and combat adjustment test scores, and to advise concerning the use of such information in the selection of personnel for lead crew and other special types of operational training.

b. To administer aptitude and/or proficiency tests when indicated for the purpose of securing additional information for the selection of personnel for special training.

c. To collect and/or develop criterion data on the proficiency of individuals in carrying out various operational duties. Such data will be used in refining present testing and selection procedures or will be forwarded to Headquarters, AAF Training Command for use in validation studies to be conducted by that Headquarters.

d. To develop new tests or test specifications designed to measure aptitudes and proficiencies that are important for success in combat and that are not measured by existing tests.

e. To undertake studies of attitudes, motivation and leadership ability as indicated by local problems or as directed by Headquarters, Army Air Forces.

Of the types of activity directed in these letters, item "c" above, on the procurement of proficiency measures and validation of test results called for the major effort on the part of research personnel.

After almost a year of research experience had been gained in the Continental Air Forces, a formal statement of the mission of aviation psychologists was published in AAF Letter No. 20-103, subject "Aviation Psychology in the Continental Air Forces," dated 19 July 1945, which may be found in the Appendix A.2. Those paragraphs relating to mission are essentially similar to the ones in the letters of 23 August 1944, but no reference was made to the development of new tests or studies of attitudes, motivation and leadership. These deletions resulted from a decision to concentrate on the other areas of research mentioned in the earlier letter.

Specific Mission of Psychological Sections in each Air Force

Some division of responsibility in the accomplishment of the over-all mission stated above was made among the psychological sections in the four air forces. Shortly after psychologists were assigned, letters with the subject "Research Project" were sent from the Office of the Air Surgeon to each air force, indicating particular lines of investigation to be followed within the framework of the general mission. In a letter dated 22 September 1944, the First Air Force unit was directed to make a follow-up study of fighter pilots in order to provide information on the relative validities of all classification tests for predicting fighter pilot success. This information was of special importance since at that time differential stanines were being assigned for fighter and bomber pilot training and an empirical check was desired on the relative weights to be given tests for the fighter pilot stanine. Suggestions as to the type of criteria to be used were made in this letter, one paragraph of which is quoted below:

> 2. It is desired that special attention be given to a follow-up study of fighter pilots assigned to the First Air Force from recent AAF Training Command Classes. This follow-up study should include collection of the records of pilots who have had accidents, who have been brought before Flying Evaluation Boards, or who have been considered unsatisfactory because of lack of proficiency or other reasons. Records of percent hits in fixed gunnery practice and records of gun camera missions should be collected. An effort should also be made to identify individuals who have been outstandingly successful in R.T.U. training.

With this assignment, research personnel in the First Air Force immediately concentrated their entire attention on fighter pilots in the First Fighter Command of that air force, in which work they continued until the close of the war. The magnitude of these studies allowed no opportu-

nity for studies with the medium and heavy bombardment crews also being trained in the First Air Force.

Letter, Headquarters AAF, 15 September 1944, subject "Research Project," directed that studies be made in the Second Air Force on problems in the formation of lead crews. This letter was written in anticipation of a subsequent interest throughout the Continental Air Forces in the selection and training of lead crews before their shipment to combat. A relevant paragraph from this letter follows:

> 2. At the present time the problem of selecting and putting together individuals for training as potential lead crews is considered to be of highest priority for the research officers in your Command. The Commanding Generals of various overseas air forces have indicated that insofar as possible, superior individuals should be identified and trained together in the Zone of the Interior so that overseas commanders can select lead crews insofar as practicable from among these potential lead crews. The development of specific procedures for identifying superior individuals has already received a great deal of attention. Available data indicate that both the stanine scores and special combat aptitude ratings computed from specific classification test scores give significant predictions of combat performance and should be interpreted and used for the purpose outlined above. It is desired that special attention be given to further work on the development of lead crew aptitude ratings.

While they were still in the Training Command, the research personnel who were assigned to the Third Air Force had been working on a study of the attitudes toward a second tour of combat duty of gunners returned from overseas. Paragraph 2 of a letter from Headquarters AAF, dated 4 October 1944, directed research personnel in this Air Force to continue similar studies on the utilization of returned pilots, bombardiers and navigators. Certain statements were made concerning other studies which should be undertaken by that research section as indicated by the succeeding paragraphs of the cited letter:

> 3. In connection with the type of study outlined in the preceding pargaraph, importance is attached to the implications of this or similar research for the more general problems of leadership, motivation and attitude in AAF personnel.
>
> 4. Research that is concerned with the problem of selecting personnel for lead crew and other special types of operational training is important as well as follow-up studies to determine the validity of present selection and classification tests for predicting proficiency in various operational duties. However, since considerable research on this problem is under way in the Second Air Force and in various overseas Air Forces, it is believed desirable at the present time for personnel in your command to give highest priority to research on the problems outlined in the preceding paragraphs.

The area of concentration for the psychological section in the Fourth Air Force was similar to that for the First, in which particular emphasis was given to studies on fighter pilots, especially those types of aircraft which were used only in the Fourth Air Force. Below is quoted a paragraph from a letter, Headquarters AAF, 6 October 1944, subject: "Research Project."

> 2. It is desired that special attention be given to a follow-up study of pilots assigned to the Fourth Air Force from recent AAF Training Command classes. In this follow-

up study special consideration should be given to night-fighter pilots and two-engine fighter pilots. Studies should be made for the purpose of determining the type of individual who is most successful as a night-fighter and a two-engine pilot and for determining the special requirements that are specific to operation of these types of planes.

This letter further indicated the desirability of obtaining additional information from instructors and commanding officers relative to pilot proficiency. Special mention was made of the importance of determining the relationship between night-vision tests and proficiency in night-fighter pilot training. It was stated that a study of night-fighter pilots should be done to yield information on the requirements of instrument flying.

Organization

At the time when psychological sections were assigned to the four air forces in September 1944, the operation of these commands was directed by Headquarters, Army Air Forces. Thus the Psychological Branch, Research Division, Office of the Air Surgeon, dealt directly with the psychological sections in the four air forces, coordinating their research work and receiving reports of progress. With the establishment of Headquarters, Continental Air Forces, a commissioned aviation psychologist was assigned to the Office of the Surgeon on 3 May 1945, and took over the detailed supervision of the psychological sections. Throughout the remainder of the war, this officer devoted his attention to bringing about a closer coordination of research on similar problems in the four air forces. With the exception of a new emphasis on the measurement of proficiency of crews as a whole, no basic change in the mission or operation of the four air forces psychological sections took place with the establishment of this intermediate headquarters section.

Psychological personnel sent to the four air forces in September 1944 were assigned to the Surgeon in each of these commands. Initially, all research personnel in the First, Third and Four Air Forces were located in these headquarters, while in the Second Air Force one officer and two enlisted men were immediately assigned to each of the following subordinate wings: 16th Bombardment Wing at Biggs Field, El Paso, Tex., where B–17 and B–29 training were carried on; 46th Bombardment Wing at Ardmore Army Air Field, Ardmore, Okla., which was engaged in heavy bombardment training using B–17's; 15th Bombardment Wing whose headquarters was located at Colorado Springs, Colo., and which was engaged in B–24 training; and 72d Fighter Wing at Peterson Field, Colorado Springs, Colo., which gave training in P–40 and P–47 fighter aircraft.

Research personnel in the Third and Fourth Air Forces remained in the headquarters organization throughout the war, while in the First Air Force, a division of personnel took place. One officer and two enlisted men were detached for a period of 90 days at Bluethenthal Field, Wilmington, North Carolina, to study long range fighter pilot training. During this period the

301st Fighter Wing requested that this psychological officer accompany the Wing overseas, which he did on an extended tour of temporary duty. A permanent division of personnel took place on 2 December 1944, at which time an operating unit was established at Richmond Army Air Base, Richmond, Va., to which all personnel, except the Chief of the Section, and the personnel on detached service, were assigned. This transfer was made because at that time this base served three functions; a basic fighter pilot training station, an in-processing base for all fighter pilots reporting to the First Air Force for training and a staging area for the out-processing of fighter pilot graduates for assignment overseas. This made it a strategic location for research operations, since psychological records could conveniently be initiated, maintained and completed for all student personnel in the First Fighter Command.

Personnel

A list of all personnel who were assigned to the psychological sections in the Continental Air Forces at any time during the existence of these sections is given in table 1.1. This roster is organized in terms of air force units and gives the name, rank, army serial number and station for each officer and enlisted man.

Specific Types of Projects Undertaken in Each Air Force

In the main body of this report an attempt will be made to present an integrated picture of the results of research in the psychological units of all four air forces, organized in terms of the findings about certain aircrew positions, and in relation to particular research problems. It is, therefore, believed appropriate to present in this section a brief account of the types of projects undertaken in each air force to show how each psychological unit went about the fulfillment of its general and specific missions.

First Air Force

Since research personnel in the First Air Force were directed to give primary attention to fighter pilots, a procedure was early established for following all student pilots through their entire training in the First Fighter Command. A basic population file was created with a card for each student containing two types of information: personnel data and proficiency measures. In the first category appeared the student's name, rank, officer serial number, name of last Training Command School and date of departure therefrom, names of each First Fighter Command Station to which assigned and dates of departure from these, stanines of all specialties, testing number and the Training Command Class number. Under proficiency measures were presented all available proficiency data. A complete file was maintained for all students entering training on or after 25 March 1944 kept up to date until the end of hostilities. The file served as a basic research tool

for studies of reliability of criteria and for validation of selection procedures. Among the studies accomplished were analyses of the errors in recording and distributions of gunnery scores, and studies of their reliabilities and of the learning displayed on successive gunnery missions. Two other types of criteria already in use at certain training bases were located and studied: mission grades and comments, and records of landings and take-offs kept by mobile control units. A great deal of attention was given to the development of certain over-all proficiency measures. Printed rating

TABLE 1.1.—*Roster of psychological research personnel in the Continental Air Forces*

Name and air force assignment	Rank	Serial Number	Station
Headquarters, Continental Air Forces			
Crawford, Meredith P.	Lt. Col.	O900365	Bolling Field, Washington, D. C.
Ward, Lewis B.	Lt. Col.	O900101	Bolling Field, Washington, D. C.
Mitchell, Philip H.	Major	O916071	Bolling Field, Washington, D. C.
Heyns, Roger W.	Capt.	O1003752	Bolling Field, Washington, D. C.
First Air Force			
Sollenberger, Richard T.	Lt. Col.	O903068	Mitchel Field, N. Y.
Henneman, Richard H.	Major	O902186	Mitchel Field, N. Y., and Richmond AAB, Va.
Hadley, Howard T.	Capt.	O565535	Mitchel Field, N. Y., and Richmond AAB, Va.
Webb, Wilse	1st Lt.	O1004540	Mitchel Field, N. Y., and Richmond AAB, Va.
Jenkins, David H.	T/Sgt.	19154048	Mitchel Field, N. Y.
Smith, Thomas W.	T/Sgt.	19081807	Mitchel Field, N. Y., and Richmond AAB, Va.
Turner, Ralph H.	T/Sgt.	35202217	Mitchel Field, N. Y., and Richmond AAB, Va.
Hausman, Howard J.	Sgt.	32355403	Mitchel Field, N. Y., and Richmond AAB, Va.
Marsh, Stewart H.	Sgt.	19160546	Mitchel Field, N. Y., and Richmond AAB, Va.
Koltun, Abraham	Sgt.	32412894	Mitchel Field, N. Y., and Richmond AAB, Va.
Second Air Force			
Ward, Lewis B.	Lt. Col.	O900101	Hq., 2AF, Colorado Springs, Colo.
Brown, Marion H.	Major	O909626	Hq., 72d Fighter Wing, Peterson Field, Colo.
Mitchell, Philip H.	Major	O916071	Hq., 16th BOT Wing, Texas and Hq., 2AF, Colorado Springs, Colo.
Killian, Frank	Capt.	O1000249	Hq., 17th BOT Wing, Sioux City AAB, Ia. and Hqs., 2AF, Colorado Springs, Colo.
Nogee, Philip	1st Lt.	O1003000	Hq., 15th BOT Wing, Colorado Springs, Colo. and Hq., 2AF, Colorado Springs, Colo.
Lawrence, Warren R.	T/Sgt.	31129606	Hq., 15th BOT Wing, Colorado Springs, Colo. and Hq., 2AF, Colorado Springs, Colo.
Glasson, Roger E.	S/Sgt.	37654094	Hq., 17th BOT Wing, Sioux City AAB, Iowa and Hq., 2AF, Colorado Springs, Colo.
Stubblefield, Eben M.	S/Sgt.	18133606	Hq., 72d BOT Wing, Peterson Field, Colo.
Sumby, William H.	S/Sgt.	31143705	Hq., 16th BOT Wing, Biggs Field, Tex. and Hq., 2AF, Colorado Springs, Colo.
Traeger, Carl	S/Sgt.	16116216	Hq., 17th BOT Wing, Sioux City AAB, Ia. and Hq., 2AF, Colorado Springs, Colo.
Begley, Joseph T.	S/Sgt.	32709293	Hq., 16th BOT Wing, Biggs Field, Texas and Hq., 2AF, Colorado Springs, Colo.
Clark, William K.	Sgt.	13077221	Hq., 72d Wing, Peterson Field, Colo.
Fletcher, Everett H.	Sgt.	3[illegible]629307	Hq., 15th BOT Wing, Colorado Springs, Colo. and Hq., 2AF, Colorado Springs, Colo.
Third Air Force			
Brown, Clarence W.	Major	O908410	Hq., 3AF, Tampa, Fla.
Lehner, George F. J.	Major	O1000257	Hq., 3AF, Tampa, Fla.
De Mott, John [illegible]	Capt.	O567881	Hq., 3AF, Tampa, Fla.
Bentz, Vernon J.	Capt.	O1821963	Hq., 3AF, Tampa, Fla.
Potter, Eugene J.	M/Sgt.	19081809	Hq., 3AF, Tampa, Fla.
Dietsch, Robert W.	T/Sgt.	35516737	Hq., 3AF, Tampa, Fla.
Jerrells, Herbert E.	S/Sgt.	19200261	Hq., 3AF, Tampa, Fla.
Oman, Herbert T. A.	S/Sgt.	38008624	Hq., 3AF, Tampa, Fla.
Rokeach, Milton	S/Sgt.	19083850	Hq., 3AF, Tampa, Fla.
Shepard, Clarence E.	S/Sgt.	35019586	Hq., 3AF, Tampa, Fla.
Petti, Henry E.	Corporal	32920069	Hq., 3AF, Tampa, Fla.
Critzer, Dorothy	Corporal	A-308783	Hq., 3AF, Tampa, Fla.
Fourth Air Force			
Ghiselli, Edwin E.	Lt. Col.	O900144	Hq., 4AF, San Francisco, Calif.
Heyns, Roger W.	Capt.	O1003752	Hq., 4AF, San Francisco, Calif.
Payne, Robert B.	1st. Lt.	O1000275	Hq., 4AF, San Francisco, Calif.
Thibaut, John W.	1st. Lt.	O1112475	Hq., 4AF, San Francisco, Calif.
Matthews, Jack	M/Sgt.	15076447	Hq., 4AF, San Francisco, Calif.
Shlien, John M.	T/Sgt.	17056652	Hq., 4AF, San Francisco, Calif.
Woodruff, Joseph L.	T/Sgt.	12091673	Hq., 4AF, San Francisco, Calif.
Whittier, William C.	S/Sgt.	19082191	Hq., 4AF, San Francisco, Calif.
Minium, Edward W.	Sgt.	39136121	Hq., 4AF, San Francisco, Calif.
Mendell, Ira C.	Sgt.	19160209	Hq., 4AF, San Francisco, Calif.
Schwartz, Ralph H.	Corporal	39038598	Hq., 4AF, San Francisco, Calif.
Thomas, Garth J.	Pvt.	37729005	Hq., 4AF, San Francisco, Calif.

scales for basic and advanced fighter pilot training were devised, analyzed and used in validation studies, and later a technique was developed for obtaining proficiency ratings from interviews with instructors. Accidents were analyzed in terms of psychological causes and were used for validation studies. A procedure was worked out to isolate groups of students of high and of low proficiency, in terms of all available criteria, to be used in extensive validation studies.

Second Air Force

Psychological research in the Second Air Force went through four phases. Preliminary exploration of the relation between aptitude scores and the degree of success in operational training was made by one officer while on temporary duty from the Training Command. Validation studies using records of Flying Evaluation Boards and proficiency ratings by instructors and directors of training on pilots, copilots, navigators and bombardiers in heavy bombardment training were done during this period. The second phase began with the arrival of other research personnel on permanent assignment, and was characterized by a detailed exploration of the types of proficiency measures available in fighter and heavy bombardment training. The reliabilities of all available criteria were studied and the usefulness of simple ratings and rankings was evaluated. During this period, in cooperation with the Operations Analysis Section of Operations and Training, plans were drawn up for the selection and training of outstanding crews as potential lead crews. These plans failed to obtain the approval of higher headquarters and were, therefore, not put into operation. In a third phase, beginning in the early part of 1945, research activities were gradually shifted from heavy to very heavy bombardment training. Time was spent in the analysis of criterion data already obtained and in a further study of the problem of selecting potential lead crews. In May 1945 a fourth change in emphasis occurred, when all personnel, other than those assigned to fighter pilot studies, worked on the problem of measuring the proficiency of B–29 crews. Descriptive rating scales were developed for the assessment of proficiency of the crew as a whole and of each of its members. These were tried out at selected OTU and CCTS bases. These latter studies represented part of a coordinated research program in the Second, Third and Fourth Air Forces on development of standard measures of crew proficiency to be used throughout the Continental Air Forces in furnishing overseas commanders with information on replacement crews, and in validating crew selection procedures being instituted at that time in the Training Command.

Third Air Force

Research personnel of the Third Air Force spent the first three months after assignment completing their survey of the qualifications of returned combat personnel for further duty. During visits made to all Third Air Force stations in connection with this work, psychologists made observa-

tions on procedures used in assessing proficiency. A second major project was undertaken in fulfillment of the section's directive to study problems in leadership. The section made two surveys of airplane commander ability and set up and operated practical procedures for the assembly of heavy bombardment (B-17) crews on the basis of compatibility. At the direction of the Commanding General, Third Air Force, a service project was completed on the utilization of negro personnel, and another survey on the disposition of aircrew trainees in the Third Air Force was done at the suggestion of the Office of the Air Surgeon. Coincident with these studies analyses were made of the relation between stanines and disposition by Flying Evaluation Boards, and also the occurrence of accidents. During the last few months of the war, Third Air Force psychologists undertook the development of specific proficiency measures for a variety of aircrew duties and participated in the general research program on the measurement of crew proficiency which was directed from Headquarters, Continental Air Forces.

Fourth Air Force

Upon assignment to Headquarters, Fourth Air Force, research personnel initiated a systematic approach to the fulfillment of the section's mission in a series of studies of P-38 fighter pilot and B-24 heavy bombardment training. These included job analyses, studies of the adequacy and reliabilities of existing criteria, validities of stanines and test scores and studies in the learning or forgetting of aircrew skills. Completion of these projects occupied most of the year. Criterion and validation studies were also done on P-61 night-fighter pilots. Assistance in the validation of tests and stanines for a large number of P-38 pilots was received from the Analysis and Records Section of AAF school of Aviation Medicine, Randolph Field, Texas. During the summer of 1945 the Fourth Air Force also participated in studies of B-29 crew and individual proficiency measurement. Research was begun at the B-29 lead crew training school at Muroc Army Air Field, Muroc, Calif., an installation operated jointly by the Fourth Air Force and the Twentieth Air Force, to which selected crews were returned from the Marianas for a period of intensive lead crew training. Assistance was rendered to operations analysts and to operations and training personnel at the School in the construction of rating scales and studies were done by the psychological section of the reliability of the ratings obtained.

Comments on the Conduct of Research Operations

In the follow-up studies made by all four air forces the general procedures were roughly the same. Personnel of each psychological section or subsection visited training bases, procured data from the squadron operations office, and elicited the cooperation of instructors and other training personnel in making ratings and discussing problems of proficiency measurement. In addition to the cooperative relations maintained with training per-

sonnel, psychologists made helpful contacts with operations analysts, statistical control sections and personnel engaged in accident analysis. Mention has already been made of certain projects carried on jointly with operations analysts. From their professional point of view, operations analysts saw and appreciated many of the problems which interested psychologists. Statistical control personnel furnished useful rosters and in some cases supplied statistical aid with IBM equipment. From accident analysis sections, psychologists obtained over-all summaries of accident data, specific accident reports and information about causes and classification of accidents. Research personnel also made useful contacts with various other sections at all echelons of command.

Certain characteristics of the training mission of the Continental Air Forces presented difficulties to the psychologists in the accomplishment of their mission. Many problems had their basis in the fact that in operational training there was little interest in differences in individual ability above a certain minimum satisfactory level of performance. This lack of interest was due largely to the commitment system which required the air forces to prepare crews and individuals for combat in the largest possible numbers. Quotas assigned made little allowance for the elimination of any individuals other than those who were obviously below minimum proficiency levels. Thus, even though training personnel recognized a considerable range in ability among students, less than three percent of all trainees were reevaluated because of poor performance. However, the small percentage of students found really unsatisfactory also reflects in part the rigid preselection which occurred in the AAF Training Command. That such selection was taking place was known to many training officers who then expected that most of the students assigned to operational training could be able to and would perform in a satisfactory manner.

As a result of the lack of emphasis upon or need for discriminating among students with regard to proficiency, few accurate records of individual performance were maintained. Research personnel obtained whatever objective scores were available including fixed gunnery, skip and dive bombing scores, circular error in bombing, records of camera and radar bombing, gun camera records and many others. The adequacy and reliability of these various measures are discussed in appropriate sections of this report.

The measurement of the proficiency of the crew as a whole, in contrast to or in addition to the assessment of the individual crew members, was not done at all by training personnel except in the research procedures initiated and supervised by the research sections in the various air forces.

Another source of problems confronting research personnel was the variation in training from one station and class to another. These variations were in amount of training (hours, missions) and in emphasis on different aspects of the curriculum. Variation occurred from one station to another and from one class to another at the same station. Direct comparisons of groups on the basis of the records were, therefore, impossible and large

scale follow-up studies were difficult. Irregularities in the amount of training received by students were caused by delays in the formation of crews when individual specialists failed to arrive on schedule at the training bases, by foreshortening of the training program to meet overseas movement requirements, and by delays due to maintenance difficulties. Irregularities from station to station in the emphasis on different parts of the curriculum had two main causes. A trend toward decentralization with delegation of responsibility for detailed training procedures to the station level characterized Continental Air Forces training, and probably resulted from the early mission of continental air force units in tactical and defense operations. This trend was somewhat reversed toward the end of the war. A second reason for differences in emphasis was the fact that instructors were drawn in increasing numbers from the ranks of aircrew members returned from overseas. Their experiences in the combat zone largely determined which phases of training they thought most important. Since they came from different theaters, their views reflected differences in tactics employed in these theaters. Thus, in completing the follow-up studies, aviation psychologists had to resort to several different procedures to correct for differences in the range and the absolute values of proficiency scores obtained from different bases.

ORGANIZATION OF THE REPORT

The basic orientation of the Aviation Psychology Program toward the individual aircrew position is reflected in the subsequent organization of this report. The first group of chapters is devoted to each aircrew specialty found in the Continental Air Forces, outlining the job requirements, methods for measurement of proficiency in the specialty, and the relation of achievement of individuals on selection and classification tests to measured performance on the job. The amount of space given to each is a function of the amount of research data accumulated during the emergency period, and does not necessarily correspond to the relative importance of that aircrew position within the crew or in the tactical formation. Following these is a chapter on the crew as a whole which discusses the problem of evaluating a crew as a working team, achieving a common goal, rather than as a collection of specialists. Subsequent chapters recount those studies which were undertaken somewhat aside from or incidental to the major objectives of the program.

CHAPTER TWO

Fighter Pilot Studies

ANALYSIS OF DUTIES

Although the detailed analysis of the fundamental job of the pilot was felt to be the responsibility of the Psychological Research Project (Pilot) in the AAF Training Command, some study of the training requirements and duties of the fighter pilot in operational training was obviously necessary. In the discussion which follows, those aspects of the requirements are stressed which are peculiar to operational training and combat. No attempt is made to present a detailed statement of training, since this varied from time to time, and also can easily be obtained from the appropriate training manuals and directives. Rather, emphasis is placed upon the results of studies made to obtain information about the qualities necessary for success, and upon those activities of the pilot which yielded criteria of proficiency.

Job Description

Information about the job of the fighter pilot in operational training was obtained from many sources: Training Standards, manuals, training directives, conferences with instructors, observations of certain aspects of training and conversations with pilots who had had combat experience. Below is given a brief outline of fighter pilot training requirements as obtained from these sources. Additional information is given under the section on criteria where more detailed descriptions are given of those operations and requirements which yielded data used as criteria of proficiency.

Combat training guides for P-47, P-40 and P-38 type aircraft were found respectively in I Fighter Command Circular 50-8 with amendments, in Second Air Force Manual 50-6 revised from time to time, and in 4th Fighter Command Memorandum 4?-8 also revised. Minimum requirements in each case included 120 hours of flight training spread over a 10- to 12-week period. In addition, each student took ground courses covering such subjects as navigation, communications, meteorology, tactics, combat intelligence, gunnery and others. The numbers and types of missions and the organization of flight training varied somewhat for the different types of aircraft and in the different commands involved. For example, with P-38 aircraft, training was at first divided into two phases taken at different stations. Later all P-38 training was done at the same stations and phase distinctions were dropped. In the Second Air Force the P-40 and P-47 training had no phase division except that all missions in aerial and ground

gunnery, rocket firing, and dive bombing were usually done at special gunnery stations to which pilots were assigned for short periods. Finally, P–47 training in the First Air Force was divided into two phases, basic and advanced, taken at different stations.

In the early part of training (basic phase) emphasis was on familiarization with the operation of the aircraft and on the simpler combat maneuvers. Experience was given in missions at different altitudes; and oxygen equipment and the anti-G suit were tried out in appropriate maneuvers. Individual and unit combat was practiced and some night flying done. In the later part of flying training (advanced phase) were included more complex maneuvers, further experience in formation flying, training in gunnery, strafing, bombing, low altitude navigation and night flying. "Combined training," involving coordinated missions with bombardment type aircraft, was an important part of this training.

In addition to the required missions, several hours of unspecified flying were allotted in the directives, allowing individual stations to supplement the required missions according to local needs. Differences in the use of this time and differences in facilities and conditions existing at different stations tended to make training less homogeneous than would be indicated by statements found in the training directives. Such variations in training added greatly to the difficulties met in obtaining adequate criteria of proficiency of the individual pilot.

Job Specifications

The factors involved in flying fighter type aircraft may be separated into two categories. They are *skills* and *traits*. This is admittedly an arbitrary division, but seems to be one which permits an efficient handling of the available data. The term skill is in general applied to a psychomotor response in a flying situation. The term trait is used in the discussion of modes of behavior which appear to spring from the psychological background of the pilot. An example of a skill is a "quick rejoin"; and an example of a trait is "aggressive in the air." The information about both the skills and the traits important for success as a fighter pilot was obtained from a number of sources. In the discussion which follows, these skills and traits are discussed separately according to source from which the information was obtained. The over-all picture provided by the traits and skills described in this section may be thought to furnish a rough set of job specifications for the successful fighter pilot.

Skills

1. *General analysis of skills.*—Aviation psychologists in the Fourth Air Force obtained information as to what were the most important flying skills for P–38 fighter pilots from four main sources of information: "Hangar Flying," or interviewing flying personnel; study of training regulations;

re-evaluation case records; and a preliminary rating scale. The skills listed below represent a psychological assessment of the information obtained from all four sources.

a. Ability to *fly close formation* without losing position as a result of lagging, falling out on turns and erratic throttle control, failure to begin maneuvers promptly, defaulting on maneuvers, failing to maintain prescribed altitude, and over-controlling.

b. Ability to *locate and fire at air targets*, estimating correctly the alignment, lead, angle and range without slipping or skidding, and using ammunition effectively.

c. Ability to *identify and fire at ground targets*, estimating correctly the proper line of flight, range and angle, coordinating controls, and controlling [illegible] effecti[illegible]

d. Ability [illegible] *perform acrobatic maneuvers* as prescribed, without defaulting o[illegible] [illegible] position, demonstrating a knowledge of the performance characteristics of the airplane.

e. [illegible]bility to *locate and bomb ground targets* with proper use of equipment and [illegible]chniques.

[illegible] *Mission card comments.*—An analysis of factors concerned with flying the P-47 was made by research personnel in First Air Force from instructor comments on mission cards. There were eight types of cards, covering the following missio[illegible] in the basic phase of training: transition, formation, navigation, instrument, acrobatics, combat maneuvers, night flying and cross-country. The comments were voluntary additions to the grading of the missions. During the course of instruction, a student pilot was graded on different missions by 6 to 10 instructors, usually several times by each. For a sample of 200 cases with a minimum of 10 cards and a minimum total of 5 comments, the analysis shown in table 2.1 was obtained. It must be said that the categories are probably to a large degree a function of the type of flying. Since the missions covered the basic flying phase, care must be exercised in drawing any general conclusions for *all* types of flying. The comments were analyzed into three categories: favorable comments, neutral remarks, and major criticisms. Distributions of the favorable and unfavorable comments for different aspects of flying are shown below. The categories are not mutually exclusive, but represent the best grouping of the comments that could logically be decided upon. It is clear from the data that instructors tended to comment favorably upon rather than criticize mission performance by a ratio of two to one. They apparently considered formation flying in general (unspec.), technique of rejoining formations in particular, and landing technique as the most important skills in operational training for fighter pilots.

3. *Deficiencies reported for pilots* [illegible]*eeting flying evaluation boards.*—Research personnel in the Fourth Air F[illegible]ce made a study of case reports of student pilots who appeared before Fl[illegible]ng Evaluation Boards because of lack of proficiency in flying P-38, P-3[illegible], and P-61 aircraft. Of 33 day-

fighter students re-evaluated, 15 were unsatisfactory in formation flying. Fourteen of the others were reported as having more general deficiencies such as poor judgment, slow reaction time, etc., some of which might be considered skills and others more nearly represented traits. Three individuals were evaluated because of recklessness or carelessness and one because of fear. A more detailed analysis of the flying deficiencies of these and three other students is given in table 2.2 which lists the frequency of reporting

TABLE 2.1.—*Relative frequencies of favorable and unfavorable comments on mission cards*

FIRST AIR FORCE

Aspect of flying	Type of comment: Favorable—a: Number	Favorable—a: Per cent	Unfavorable—b: Number	Unfavorable—b: Per cent	Ratio: A-B
Skills					
Formation flying:					
Formation (unspec.)	714	26.8	156	14.8	4.58
Rejoin	266	10.0	243	23.0	1.09
Position	285	10.7	70	6.6	4.07
Holding altitude	38	1.4	44	4.2	.86
Use of turbo	58	2.2	6	.6	9.67
Smoothness on controls	2	.1	31	2.9	.06
Combined	1,363	41.2	550	52.1	2.48
Basic flying technique:					
Landing	269	10.1	199	18.9	1.35
Take-off	176	6.6	96	9.1	1.83
Traffic pattern	71	2.7	42	4.0	1.69
Combined	516	19.4	337	31.9	1.53
Special skills:					
Instruments	184	6.9	17	1.6	10.82
Navigation	136	5.1	33	3.1	4.12
Radio procedure	60	2.3	31	2.9	1.96
Combined	380	14.3	81	7.7	4.69
Combat maneuvers	152	5.7	24	2.3	6.33
Acrobatics	128	4.8	16	1.5	8.00
Total (skills)	2,539	95.4	1,008	95.5	2.52
Traits					
Judgment	67	2.5	23	2.2	2.91
Air discipline	51	1.9	21	2.0	2.43
Cockpit procedure	4	0.2	3	.3	1.33
Total (traits)	122	4.6	47	4.5	2.60
Grand total	2,661	100.0	1,055	100.0	2.52

of each of several categories of deficiencies. The data show several deficiencies were reported for some of these student pilots. Deficiencies in skills were apparently reported about twice as frequently as trait deficiencies in the student pilots considered for re-evaluation. In the discussion under traits, it will be shown that many of the observed deficiencies in skills were actually thought by instructors and flight surgeons to be caused by trait deficiencies. Hence, no great weight should be given to the relative frequencies with which traits and skills were given as official reasons for re-evaluation.

Since formation flying was so frequently reported as a reason for re-

evaluation, a special analysis was made of deficiencies reported in this category. The results are given in table 2.3.

Finally, the principal flying deficiencies resulting in Flying Evaluation Board Action were examined for 13 re-evaluated night-fighter students. All but three of these were reported as failing to meet night-fighter instrument standards, probably largely a matter of deficiencies in skills. One each was re-evaluated for over-cautiousness, lack of confidence, and "poor pilot technique."

TABLE 2.2.—*Specific deficiencies reported for thirty-six students considered for re-evaluation*

FOURTH AIR FORCE

Deficiencies	*Frequency*
Skills:	
Specific flying situations:	
Formation flying	22
Acrobatics	5
Gunnery patterns	2
Landings	2
Total	31
Flying technique:	
Slow progress (slow learning)	8
Slow reaction time	7
"Below minimum standards"	4
Poor coordination	3
Roughness in control	1
Overcontrol	1
Total	24
Total skills	55
Traits:	
Dangerous, reckless, careless	8
Poor judgment	6
Lacks discipline	4
Lacks confidence	4
Too cautious and deliberate	3
Lacks aggressiveness	2
Poor attitude toward his duties	1
Total traits	28

4. *Student errors noted by mobile control unit observers.*—In the First and Second Air Forces violations noted by the observer in the mobile control units were used as a criterion of pilot proficiency. Research personnel in the First Air Force selected as a sample the violations of 100 students over a 30-day period for an analysis of landing and take-off errors. As might be expected, there are some cases of overlapping in the categories used. For instance, in landing it was usual that the approach was low if the base leg and final turn were low. Also if the approach was "hot," the landing was most apt to be "hot." In some cases, the decision as to where the

error was reported was based on whether or not the pilot actually landed or was instructed to go around. For example, the former might be reported as a "hot" landing, while the latter was reported a "hot" approach. In table 2.4 is presented the distribution obtained.

Traits

1. *General analysis of traits.*—An obvious source of information about traits important for success as a fighter pilot was interviews with instructors and fighter pilots with combat experience. Research personnel of both

TABLE 2.3.—*Frequency distribution of criticisms made of the formation flying of twenty-two re-evaluated students*

FOURTH AIR FORCE

Criticism	*Frequency*
Failure to maintain position on turns:	
Falls out on turns	7
Lags on turns	2
Loses altitude on turns	2
Only shallow turns	1
Total	12
Failure to maintain position on maneuvers other than turns:	
Slow to join formation	4
Starts maneuvers late	3
Lags in rat race	1
Poor acrobatics	1
Lags on take-off	1
Total	10
Failure to maintain position in straight and level flight:	
Straggling, lagging	5
Erratic	5
Failure to fly close	1
Flies below	1
Total	12
General criticisms:	
Dangerous	5
Slow reaction time	1
Total	6

the First and Fourth Air Forces obtained considerable information from this source. In addition, training manuals, rating procedures, and the like, yielded information from which inferences could be made as to the importance of various traits. The following is a list of the most important of the traits identified by research personnel of one or both Air Forces, together with phrases offering brief trait descriptions in each case.

a. *Alertness.*—Keeps continually alert while flying, spots installations and other aircraft quickly, makes quick, intelligent responses to instructions, orders, signals, evidences of malfunction in the aircraft.

b. *Friendliness—Cooperation.*—Is cheerful and agreeable, friendly with everyone, generally well liked, "wears well," is willing to sacrifice individual desires in the interests of combat teamwork, shows no harmful individualism, is able to work with others routinely without friction.

c. *Eagerness for combat.*—Has strong desire for action with plane against the enemy, as student is prompt and enthusiastic about the routine as well as more adventurous tasks, likes to be on offensive in simulated combat maneuvers, is actively and positively aware of combat flying as the goal of operational training.

TABLE 2.4.—*Errors made by 100 student pilots and noted by mobile control observers*

FIRST AIR FORCE

Flight description	*Number of errors*
Base leg:	
Wide, far out	31
Low	14
Total	45
Final turn: Low	23
Approach:	
Low	74
Dragged in	40
Spacing too close	26
Hot	23
High, overshot	17
Long	5
Slipping	5
Total	190
Landing:	
Long	29
Hot	20
Short	8
Rough (dropped in)	6
Wheel	5
Total	68
Other: Wheels down late	2
Take-off:	
Low turn	33
Long (tail high)	5
Total	38
Total number of errors	366

d. *Leadership.*—Accepts and discharges responsibility well, is willing to take the lead in new situations, is respected by other students.

e. *Air discipline.*—Executes orders promptly, obeys instructions and follows flying regulations, thinks and flies for the success of the whole group regardless of individual considerations.

f. *Dependability.*—Is dependable, has habitual attitude of responsibility, behavior is stable and predictable.

g. *Endurance.*—Has physical and moral stamina to stand long hours of flying.

h. *Industriousness.*—Devotes time and energy to training beyond formal requirements, is determined to succeed, puts into each job more than is required, extends himself to learn new flying procedures, enemy tactics, and the limitations and possibilities of his plane.

i. *Maturity of judgment.*—Avoids hazardous, expensive practices, avoids thoughtless violations of regulations resulting in inefficient use of aircraft, follows advice of experienced pilots.

j. *Attention to detail.*—Has a wholesome respect for detailed procedures developed for the safety of flying personnel and follows them exactly.

k. *Courage.*—Is able to fly close formation without defaulting on dangerous maneuvers, has few abortive missions.

l. *Military discipline.*—Has respect for superiors, obeys regulations and orders, is prompt, is neat, etc.

m. *Possibility of growth.*—Is able to develop as new tasks are given, grasps new problems, learns new lessons and accepts increasing responsibility.

n. *Morale.*—Makes the best of the situation in which he finds himself, has no interfering personal worries, will not let combat losses get him down.

Obviously, some of the above traits considerably overlap others. It is also likely that there would be disagreement among observers as to the importance of many of them. For traits identified in the First Air Force, an attempt was made to obtain evidence of both discreteness and overlapping. From interviews with combat pilots, 12 traits had been identified as differentiating the better combat pilots from the poorer. These traits were then ranked by the experienced pilots in the order of their relative significance. Comments were also requested regarding the discreteness of the traits and clarity of definition. The results are summarized in table 2.5.

TABLE 2.5.—*Evaluation of traits significant for fighter pilots*

FIRST AIR FORCE

(N–13)

Trait	Mean Rank	Relative Rank	Discreteness[1]
Air discipline and teamwork	2.73	1	0
Eagerness for combat	3.69	2	0
Leadership	[2]3.82	(2) 3	2
Aggressiveness	5.11	4	4
Decisiveness	5.54	5	7
Coolness	5.92	6	3
Conscientiousness	6.23	7	1
Alertness	6.96	8	3
Morale	[2]7.71	(1) 9	4
Tough thinking	[2]8.00	(1) 10	4
Endurance	9.15	11	0
Personality	9.69	12	0

[1] Number of pilots who stated this trait to be the same as, or a part of, another trait listed.

[2] Refers to the number of pilots who stated this trait to be a summation of all the other traits listed.

On the basis of ratings of discreteness shown in the table, the traits of aggressiveness, decisiveness, and tough thinking were combined with each other or with other traits in the list. A similar procedure applied to the more extensive list from both Air Forces would probably result in additional consolidation and elimination.

2. *Comments on mission cards.*—Since less than 5 percent of the mission card comments referred to traits, the particular traits mentioned probably have little significance. "Judgment," "air discipline" and "cockpit procedure" seemed to be the only traits directly observed by the instructors grading mission performance. This is perhaps only natural since a trait is ordinarily thought an over-all continuing characteristic to be inferred from many observations. Mission comments usually referred to specific observations of behavior on a single mission.

3. *Psychological causations of pilot-error accidents.*—Another method of obtaining information about qualities important for pilot success is to study the qualities of individuals who fail, particularly those qualities that might have contributed to the failure. In a sense, pilots that have aircraft accidents due to pilot error may be considered to have failed in one aspect of their job. Hence, a study of the psychological reasons for such accidents would seem to be worth while.

An analysis of P–47 pilot-error accidents in terms of probable psychological causation was made by the First Air Force research personnel from the data recorded on the accident reports sent in to Headquarters, I Fighter Command. The examination of the individual accident records consisted of two steps. The first involved abstracting all pertinent information from the records on each case. When all records had been abstracted, each case was taken up for analysis by a committee of three to five men. Each member, working independently, made his own assessment of probable psychological causation from the information on the accident in the abstracted record. Following this step, differences in assessment of a psychological trait were discussed and eliminated in the committee. A summary table of the psychological categories and their relative frequencies is presented in table 2.6. A similar procedure yielded somewhat different results in a study of 100 consecutive P–47 accidents in the Second Air Force. Data from this analysis are also shown in table 2.6. The relative frequencies of different categories of reasons in the two distributions are somewhat different. Either the nature of the accidents was different in the two commands or the psychologists differed markedly in their interpretation of the reports.

4. *Reasons for flying evaluation board action.*—Studies of reasons given for re-evaluation of fighter pilots by Flying Evaluation Boards in the Fourth Air Force indicated that traits were reported approximately one-half as often as skills. (*See* table 2.3, page 8). However, in the reports *traits* were frequently listed as explanations or reasons for the lack of *skill* reported. In table 2.7 are given the explanations in terms of traits offered by flight surgeons and instructors for deficiencies in skill that were reported. Thus,

although more than half of the re-evaluations were ascribed to lack of skill, it was apparently believed unfavorable traits were fundamentally responsible in most cases.

5. *Study of traits on the Minnesota personality inventory in relation to accidents and flying evaluation board action.*—As a cooperative project with the psychiatrist at the P-47 Fighter Pilot Wing Indoctrination Unit in the Second Air Force, the Minnesota Multiphasic Personality Inventory was administered to all incoming trainees and combat returnees in the fall of 1944. Data are available for this population on the incidence of aircraft accidents and elimination from training by FEB action. It was thought that

TABLE 2.6.—*Psychological causes contributing to pilot-error accidents*

CATEGORY OF PSYCHOLOGICAL CAUSATION

First Air Force Study

	N	*Percent*
Speed-distance miscalculation	41	26
Incautious planning	36	23
Poor division of attention	35	23
Reaction after imposed emergency	27	17
a. Speed-distance miscalculation	(9)	(6)
b. Poor handling of airplane	(18)	(12)
Poor plant judgment	9	6
Recklessness	7	5
Total	155	100

Second Air Force Study

	N	*Percent*	*Corresponding percent 1AF*
Poor judgment (miscalculation of speed, etc.)	14	17	32
Incautious planning (includes recklessness)	41	51	28
Reaction after imposed emergency	8	10	17
Poor division of attention	17	21	23
Miscellaneous, not classifiable	1	1	0
Total	81	100	100

a comparison of the test scores of pilots who later had accidents or were eliminated from training with scores of pilots who were more successful in training might reveal personality differences that might be of predictive value. In table 2.8 are presented the mean scores for pilots with aircraft accidents and/or who were eliminated by FEB action and mean scores for pilots more successful in operational training. Except for a larger percentage of doubtful answers among the failure group, none of the differences between the two groups is of significance. The trait where the difference most nearly approaches significance (8 percent level or less) is Pt or "Psychasthenia" and here the failure group shows the adverse trait even less than the success group. Since the lie score is also slightly greater for the failure group, no great significance can be attached to any of the observed differences.

Of perhaps more interest is the generally favorable score on most traits of the total group. The main exception here is the poor or "abnormal" score on Ma, Hypomania. Evidently the fighter pilot differs from the norm of the general population rather markedly in this trait. This coincides with the popular view of the ideal fighter pilot as slightly manic and overactive.

TABLE 2.7.—*Flight instructors' and flight surgeons' explanation for the Specific deficiencies in re-evaluated students*

FOURTH AIR FORCE

Explanation	*Number of times mentioned*
Personality and cha: icter defects:	
Defects not directly associated with flying:	
Emotionally immature	3
Emotionally Unstable	2
Lack of discipline	2
Carelessness	1
Total	8
Defects directly associated with flying:	
Fear	14
Lack of confi'lence	7
"Temperamentally unsuited"	6
Nonaggressive	4
Psychiatric conditions	3
No desire to learn	3
Tenseness	2
Dislike for acrobatics	2
Total	41
Defects in skill:	
General defects ("la:k of ap'itude")	8
Poor judgment	5
Slow reaction time	5
Slow learning rate	2
Total	20

CRIT RIA OF PROFICIENCY

The types of mi:sions a :d main emphasis of training have been described in the previous section. 'ot all of the missions or all of the training activities were suitable for ı se in evaluating pilot proficiency. Certain missions and activities yielc i relatively more useful measures than others. Many provided no inforr ı tion whatever except number of hours of flying accomplished. In the fc ɔwing discussions of the individual criteria an attempt will be made t evaluate each in terms of certain standards of usefulness. The main co iderations are: th nature of the distributions of scores (dispersion and 'ewness); the reliability of the scores (internal consistency); the capacity to discriminate between individuals known to

differ in proficiency; the extent to which all aspects of the job are covered; the objectivity (susceptibility to personal bias, general impression errors, etc.); and susceptibility to miscellaneous and special sources of error. Finally, at the end of the section on Criteria, there is an over-all evaluation, including comparisons between various different criteria.

Single Criteria

Fixed Gunnery, Bombing and Rocket Firing

1. *Description.*—In the advanced phase or latter half of fighter pilot training at the operational level, the principal emphasis was upon instruction in gunnery, bombing, and rocket firing. Minimum requirements were the firing of 2,000 rounds per student for P–38, P–40 and P–47 students.

TABLE 2.8.—*Comparison of scores on the Minnesota multiphasic inventory*

163 Pilots having aircraft accidents or FEB action

804 Pilots without accidents or FEB elimination

SECOND AIR FORCE

Trait description	Code	Failure Mean	Group S. D.	Success Mean	Group S. D.
Doubtful Answers	?	49.3	2.3	50.3	2.6
Lie Score	Lie	52.3	4.0	51.7	3.6
Bizarre Answers	F	51.1	2.6	51.4	3.4
Hypochondriasis	Hs	43.7	4.5	43.8	4.7
Depression	D	45.0	7.5	44.4	7.6
Hysteria	Hy	52.6	7.0	51.6	6.8
Psychopathy	Pd	50.0	9.3	49.3	8.7
Masculinity	Mf	48.9	9.0	49.9	9.0
Paranoia	Pa	50.3	6.4	50.3	7.2
Psychasthenia	Pt	44.2	6.3	45.8	7.4
Schizophrenia	Sc	45.1	5.6	46.1	6.6
Hypomania	Ma	57.0	9.5	56.6	9.4

A maximum of 3,000 to 3,600 rounds was permitted. Gun camera missions up to a maximum of 50 percent of aerial gunnery training *time* were permitted, provided the minimum rounds-fired requirement was met.

Gunnery training missions were of three types: camera, air-to-air firing, and air-to-ground firing. Air-to-air firing was done at low, medium, and high altitudes. The target was towed approximately 1,000 feet behind a TB-26 flying straight and level at 175 to 180 m. p. h. Scores on high altitude missions were usually low and such missions were seldom scored, being flown for familiarization only. The data below on air-to-air firing were taken from low and medium altitude missions. The usual target was 6 by 30 feet in size (180 sq. ft.). The number of hits on the target was divided by the number of rounds fired to provide the "gunnery score" (i.e., percent hits). Because of shortages of materials, targets for air-to-air firing were not always of the same type or size, although attempts were made to compensate for this variation by converting the percent hits obtained to equivalent scores on a standard target. Whenever an atypical or frayed target was used, a corrected hits score was obtained by the formula: Corrected score = (180/actual target area) × (actual hits). Records of

scorable rounds fired and numbers of hits were maintained by the gunnery officers or by record clerks in the training sections.

In ground gunnery the training directives specified that a minimum of 1000 rounds be fired for record at ground targets 36 square feet in area. This ground gunnery was considered a "build up and preliminary for aerial gunnery and strafing." Since a pilot usually obtained quite a few hits on the target, it gave him confidence in himself and his plane. Bombing training included both dive and skip bombing missions. The minimum requirements in skip bombing were 12 bombs dropped singly upon a vertical target 10 by 20 feet. In dive bombing students were required to drop 16 bombs singly upon a target 300 feet in diameter for P–38 aircraft and 100 feet in diameter for P–40 and P–47 aircraft. The bombing was usually scheduled in conjunction with other missions and was supposed to teach proper approach technique, importance of correct trim, use of the gun sight, and in dive bombing how to break downward at the same point as the leader. Rocket firing was introduced into fighter pilot training in 1945, but no data on rocket proficiency have been reported by the research units in the Continental Air Forces.

2. *Disadvantages as a criterion.*—Use of gunnery scores as student pilot proficiency data encountered a number of characteristic difficulties which have plagued research workers in this field. Among these were the following:

a. Training procedures and requirements were in process of being changed almost continually during the fall, winter, and spring of 1944–45, rendering not comparable the data obtained on successive classes of student pilots.

b. Station differences in recording gunnery scores, in scheduling missions, in providing incentives and in instructional procedures led also to lack of comparability of scores from station to station, even within a single air force.

c. The number of scorable missions (especially for air-to-air gunnery) varied widely from student to student even in the same station and class. These discrepancies were attributable both to differences in number of missions flown and to such factors as number of lost targets, gun jams, etc., which determined the number of scoreable missions out of those actually flown.

d. The number of rounds fired varied greatly from student to student both on single missions and for the entire course of training.

e. The conversion formula designed to compensate for different sized targets probably did not yield strictly comparable scores in all cases.

These difficulties led to extensive differences from sample to sample of student pilots (by station and class) and obviously acted as spurious factors affecting the reliability of the gunnery scores. A serious implication of these differences was the necessity of calculating reliabilities and stanine validities separately for each sample of pilots obtained. Furthermore, there

was introduced the problem of the best method to employ in combining the various small samples in order to achieve statistics based on a sufficient number of cases to warrant safe generalizations.

3. *Factors influencing gunnery scores.*—Aviation psychologists in the First Air Force made a systematic attempt to investigate the factors determining the gunnery scores of student pilots. This personnel spent several weeks at advanced training stations of the I Fighter Command observing firsthand the gunnery training process. The investigation revealed many factors affecting gunnery scores, roughly falling into seven categories as follows:

a. *Training methods.*—proper briefing and critique, number of camera missions before start of firing, number and type of missions flown per day, time elapsing between gunnery-type missions, and time elapsing between a gunnery mission and its assessment.

b. *Motivation and interest.*—Several techniques were employed; knowledge of success and comparative standing in the group, rewards in the form of cross-country flights and early leaves of absence, penalties in being held back for failure to qualify in gunnery, etc. Extensive differences were found among stations in the use of incentives. Those stations emphasizing motivation the most, consistently reported the highest gunnery scores.

c. *Instructors.*—Positive correlation (e.g., coefficient of 0.67 on 1 sample of 33 instructors and 60 students, based on 14 or more instructor missions) were found between total percent hits of student pilots and the average percent hits of individual instructors. It was further found that for class after class, student pilots at those stations where instructors' gunnery scores were low did less well than students at stations where instructors' gunnery scores were higher. The latter of course may or may not have been the result of a tendency for students and their instructors to have similar gunnery skill. Station differences in scoring, gunnery procedures, etc., may have influenced the scores of both instructors and students.

d. *Mechanical factors.*—Principal factors include harmonization of the guns, condition of the gun mounts, malfunctions, size of target and conversion factor for targets of different size, individual aircraft characteristics and shot-away targets. Of these, harmonization and size of the target were generally considered to be the most important.

e. *Weather.*—Wind and visibility are the weather components most affecting gunnery scores. Both day-to-day and seasonal variations probably influence gunnery performance. Quantification of weather factors to provide for statistical analysis is difficult.

f. *Scoring procedure.*—Variable factors in this category are the following: identification of hits on the target, conversion factor to equate for different sizes of targets, errors in recording hits and computing percentages, and errors in counting the number of unfired rounds at the end of the mission.

g. *Flying experience.*—Data summarized below indicate clearly that

among student pilots those with more experience (e.g., former instructors) consistently shoot more accurately than do newly rated pilots.

It was concluded that in addition to the obvious factor of proficiency, the factors of motivation, effectiveness of instruction, weather and scoring procedure (in air-to-air gunnery only) were the most important sources of individual differences in the scores attained.

4. *Distribution of scores.*—The means and standard deviations of gunnery scores for different samples of student pilots varied as much within a single air force as between different air forces. For example, for 10 classes at each of three stations in the First Air Force the mean air-to-air gunnery scores varied from 2.07 to 12.55. The standard deviations for these data varied from 1.44 to 6.77. The number of pilots per class varied from 16 to 53, the median size being 40. Similar data from several classes at each of 4 stations in the Fourth Air Force gave mean air-to-air gunnery scores ranging from 1.96 to 8.74 and class size varied from 27 to 131 with a median of 70. Standard deviations ranged from 1.40 to 4.11. Corresponding data from the Second Air Force may be found in table 2.11 presenting reliability coefficients in the next section of this chapter.

Air-to-ground gunnery scores gave similar results with means varying from 6.71 to 21.55 in the First Air Force and 4.63 to 21.03 in the Fourth Air Force. Corresponding variations in standard deviations were 2.31 to 9.63 and 3.04 to 11.94 in the First and Fourth Air Forces respectively. Data on dive bombing were available only in the Fourth Air Force where mean scores ranged from 22.56 to 82.96 with standard deviations from 3.03 to 28.74.

In the spring of 1945, the First and Second Air Forces began to place much greater emphasis upon gunnery and bombing scores, setting up minimum qualifying scores and stressing improved methods of training. The consequent improvement in gunnery scores, especially in air-to-air gunnery, is strikingly demonstrated in table 2.11 for data from the Second Air Force. The data from the First Air Force showed a similar trend since the highest mean scores were obtained with the most recent classes at each station.

Aviation psychologists in the Fourth Air Force made a study of the shapes of the distributions of scores of P–38 pilots in air-to-air and air-to-ground gunnery and in dive and skip bombing. Raw scores for each station and class were transformed into standard scores with a mean of 50 and a sigma of 10. The data are presented in table 2.9. Frequency distributions of air and ground gunnery scores obtained in the First and Second Air Forces were essentially similar to those shown in the table. The distributions of scores in air-to-air gunnery, (and also in air-to-ground gunnery, but to a lesser extent), are markedly skewed, since most of the pilots made scores near the poor end of the range of scores. The reason this particular type of distribution was obtained may lie partly in the fact that pilots could not get less than zero hits and partly in the fact that an attempt was made in training to make each pilot attain at least a minimum level of

performance. Distributions of scores affected by a cut-off point or limit of performance are likely to show a piling up of frequencies near such a point or limiting score. In contrast to the gunnery scores, those for dive and skip bombing were fairly normally distributed, those for skip bombing, however, being slightly skewed in the opposite direction to gunnery scores. These permitted a relatively continuous distribution of amount of error, since few pilots approached zero scores. The differences in shape of these distributions may be an artifact of the scoring system.

5. *Reliabilities.*—Reliability studies on fixed gunnery scores have been

TABLE 2.9.—*Frequency distributions (in percents) of four proficiency measures on P-38 pilots*

FOURTH AIR FORCE

Standard score	Air gunnery	Ground gunnery	Dive bombing	Skip bombing
	Percent	*Percent*	*Percent*	*Percent*
100	0.3			
95	.2			
90	.2	0.2		
85	.6	.3		
80	1.0	.9	0.1	
75	1.8	1.5	.4	0.2
70	2.3	2.0	1.5	1.6
65	4.5	6.0	8.1	7.4
60	6.8	9.0	11.9	13.5
55	13.3	13.6	16.1	17.4
50	20.8	21.0	18.1	21.9
45	24.2	21.8	20.1	16.4
40	20.7	17.1	11.9	9.3
35	3.7	4.9	6.9	5.5
30		.8	3.7	3.6
25			.8	1.7
20			.2	.0
15				.2
Total	100.0	100.0	100.0	100.0
Number of cases	1,597	1,204	839	471

reported from the First, Second, and Fourth Air Forces. The First Air Force reported two studies differing in method. Both studies involved the correlation of scores on odd and even missions. The first study was based on the same number of missions for all pilots in a given sample, which meant throwing away scores on later missions for those men whose records showed a larger number of scoreable missions. Since many pilots had only four to six scored missions, only this small number of missions was utilized in the study. Hence the size of the correlation coefficients to be expected was necessarily limited. The results of this study are summarized in table 2.10.

In the second reliability study from the First Air Force an attempt was made to utilize all of the data available, including the scores on later missions. The students within a single station were combined and regrouped on the basis of the number of scoreable missions on their records. The odd and even mission scores were then correlated for each of these subgroups (i.e., four-mission group, six-mission group, eight-mission group, etc.) in-

dependently, the resulting correlations being later combined by Fisher's *z*. The results for three training stations are summarized in table 2.10.

For the data obtained in the Second Air Force, the correlations between

TABLE 2.10.—*Reliability coefficients of gunnery score for P-47 pilots*

FIRST AIR FORCE

4-6 Missions Only

Air-to-Air Gunnery

Station	*N*	*r* (uncorrected)
Bradley	77	.20
Bluethenthal	98	.09
Millville	44	.34
Norfolk	81	.59
Dover	86	.44
Average *r* (by Fisher's *z*)	386	.34
Estimated Reliability	...	.51
Air-to-Ground Gunnery		
Bradley	89	.25
Bluethenthal	95	.32
Millville	165	.29
Average *r* (by Fisher's *z*)	349	.29
Estimated Reliability	...	.45
All Missions		
Air-to-Air Gunnery		
Dover	184	.31
Millville	135	.38
Norfolk	144	.50
Average (by Fisher's *z*)	463	.39
Estimated Reliability	...	.56
Air-to-Ground Gunnery		
Dover	175	.45
Millville	91	.15
Norfolk	139	.38
Average (by Fisher's *z*)	405	.36
Estimated Reliability	...	.53

gunnery scores for odd and even missions (based on 14 missions) are given in table 2.11.

Tables 2.12 and 2.13 summarize the reliability studies reported from the Fourth Air Force for gunnery and bombing scores. The coefficients reported are based on the correlation between scores made on odd and even missions, all scoreable missions being included in the computations. The *r*'s for the individual samples in these tables have been adjusted by the Spearman-Brown formula. The weighted averages of the corrected *r*'s given in the tables are thus comparable to the estimated total reliabilities reported from the First and Second Air Forces. Separate data are given for

P–38 and P–61 pilots; and there was a marked difference in the reliabilities obtained for the two types of aircraft. A possible explanation of the relatively low degree of reliability found for the air-to-air gunnery scores of P–38 pilots is the fact that P–38 pilots did their firing in several different

TABLE 2.11.—*Reliability coefficients of gunnery scores for P–47 pilots*

AIR-TO-AIR GUNNERY—SECOND AIR FORCE

December, January, and February

Station	N	Odd missions Mean	Odd missions S. D.	Even missions Mean	Even missions S. D.	r (uncorrected)
Abilene	56	5.32	4.8	8.09	5.7	0.65
Ft. Sumner	38	5.55	3.2	5.32	3.6	.45
Strother	32	6.61	4.0	6.54	3.4	.30
Sweetwater	27	8.40	5.6	8.35	6.3	.83
Greenville	23	5.73	3.2	5.36	3.5	.35
Peterson	77	2.99	1.9	2.89	1.8	.36
Average	[1]253					.52
Estimated reliability						0.68
March, April, and May						
Greenville	29	17.9	8.44	15.7	7.68	0.76
Strother	42	19.9	6.77	12.3	5.33	.77
Dalhart	28	13.1	5.87	14.2	6.37	.59
Sweetwater	24	17.4	6.94	16.9	7.23	.72
Ft. Sumner	52	15.5	6.77	14.4	5.51	.62
Abilene	71	13.3	7.09	12.3	6.23	.66
Average	[1]246					.68
Estimated reliability						0.81

[1] By Fisher's z.

TABLE 2.12.—*Reliability coefficients of gunnery and bombing scores for P–38 pilots*

FOURTH AIR FORCE

Station and commitment	Air-to-air gunnery N	Air-to-air gunnery r (corrected)[1]	Air-to-ground gunnery N	Air-to-ground gunnery r (corrected)[1]
VNIII-1	68	0.00		
VNIII-2	72	.35		
SMIII-1	42	.27	42	0.43
SMIII-2	48	.56	49	.36
PVI	63	.00	50	.29
PV	58	.53		
EA	39	.22	37	.54
EB	37	.70	38	.55
EC	35	.45		
ED	36	.75		
OII	76	.28		
OIII	139	.28	57	.69
OIV	137	.00		
OV	108	.61	84	.21
Weighted average	958	.31	357	.41
	Dive bombing		**Skip bombing**	
VNIII-1	65	0.48		
VNIII-2	56	.47		
SMIII-1	28	.28		
SMIII-2	40	.49		
PV	70	0	70	0.40
PVI	55	.26	55	.29
E			41	.42
Weighted average	314	.41	166	.37

[1] Corrected by Spearman-Brown formula for each sample separately.

models of the aircraft. Since it was not possible to control this factor in the computations of reliability, the obtained coefficients may have been somewhat lowered thereby.

Table 2.14 contains a summary of the average reliability coefficients for air-to-air gunnery reported by the First, Second and Fourth Air Forces.

6. *Evaluation.*—One important consideration in estimating the adequacy of a criterion of proficiency is the extent to which it reflects known differences in proficiency. A study of this type was accomplished in the First Air Force where a comparison was made of the gunnery scores of pilots with varying degrees of experience. The hypothesis was that if gunnery scores were a valid criterion of fighter pilot proficiency, the more experienced pilots, particularly those with P–47 experience, should make higher average scores. The study was based on the gunnery records of pilots completing training during the period from January through April 1945 in the I Fighter Command. Four levels of experience were studied: pilots recently graduated from the AAF Training Command; pilots who had previously been instructors in the AAF Training Command; pilots who had previously been P–47 instructors in basic training at stations in the I Fighter Command; and pilots who had previously been P–47 instructors at advanced training stations of that Command. The average scores for these four groups are shown in table 2.15.

TABLE 2.13.—*Reliability coefficients of gunnery scores for P–61 pilots*

FOURTH AIR FORCE

Criterion	N	r (corrected)
Air-to-air gunnery	77	0.77
Air-to-ground gunnery	67	.92

TABLE 2.14.—*Average reliability coefficients for air-to-air gunnery scores*

FIRST, SECOND, AND FOURTH AIR FORCES

Air force	Aircraft	Total N	Average r (corrected)
First (first study)	P-47	386	0.51
First (second study)	P-47	463	.56
Second (first study)	P-47	253	.68
Second (second study)	P-47	246	.81
Fourth (first study)	P-38	958	.31
Fourth (second study)	P-61	77	.77

The average scores for aerial gunnery indicate there was a consistent positive correspondence between level of experience and excellence of performance. This relation is not quite so clear for ground gunnery scores, but in all cases advanced instructors had the highest scores and new trainee pilots had the lowest. Since the number of P–47 instructors (from both basic and advanced training stations) was so small, no attempt was made to calculate critical ratios between their average scores and those of pilots

in the other categories of experience. However, scores were available for larger numbers of former AAF Training Command instructors and newly rated pilots and the statistical significance of differences between the average scores for these categories is given in table 2.16. In the Second Air Force, the correlation between average gunnery scores of P–40 pilots and number of hours of previous P–40 flying time was found to be 0.68 for one class of 35 trainees.

The differences in First Air Force data favoring the experienced pilots

TABLE 2.15.—*A comparison of the gunnery scores of P–47 pilots of four levels of experience*

FIRST AIR FORCE

Air-to-Air Gunnery

Station	Advanced instructors	Basic instructors	Training command instructors	Newly rated pilots
Norfolk:				
N	19	10	40	156
Mean	11.9	9.0	8.6	7.6
S. D.	5.6	7.9	3.9	3.5
Millville:				
N	13	37	42	126
Mean	5.6	4.4	3.9	3.6
S. D.	2.8	2.7	2.2	2.8
Suffolk:				
N	....	....	44	50
Mean	....	....	6.7	4.7
S. D.	....	....	4.1	3.0

Air-to-Ground Gunnery

Station	Advanced instructors	Basic instructors	Training command instructors	Newly rated pilots
Norfolk:				
N	20	6	40	155
Mean	24.1	13.3	16.0	12.9
S. D.	8.3	5.3	9.0	5.0
Millville:				
N	13	37	41	125
Mean	22.2	15.3	18.8	15.0
S. D.	12.0	7.8	8.5	7.2
Suffolk:				
N	....	....	43	50
Mean	....	....	13.2	13.8
S. D.	....	....	7.2	5.1

and the correlation between gunnery score and hours of experience in the Second Air Force data appear to warrant the conclusion that gunnery scores do differentiate between the performance of experts (e.g., former instructors with many hours of flying time) and that of novices (e.g., recently rated pilots with relatively little flying time).

Although some of the reliability estimates appearing in the tables in this section are of somewhat lower magnitude than that generally desired for validation studies, the fact that the *r*'s on the individual small samples are so consistently positive is indicative of a degree of reliability sufficient to warrant the use of these criteria as measures of individual differences in pilot proficiency. As was pointed out earlier, the necessarily small numbers of missions upon which some of the studies were based, imposed some

limits on the size of the reliability coefficients which could be expected. And there is also the known source of error in the model of plane used in the case of the lowest reliability reported, that of 0.31 for the P-38 pilots.

In summary, it should be pointed out: a. Gunnery, bombing and rocket-firing scores are relatively objective and are little affected by personal bias of any sort. Such errors as were noted could probably be controlled by minor changes in training procedures. Many early sources of error were in process of being overcome at the close of hostilities. b. These scores have a logical validity because of similarity to the ultimate combat objectives. c. They reflect known differences in proficiency. d. And finally, they have a moderate degree of reliability. Gunnery scores might thus be expected to provide a useful criterion for validation of selection tests and procedures.

TABLE 2.16.—*Mean differences in gunnery scores for AAF Training Command instructors and newly rated pilots*

I FIGHTER COMMAND ADVANCED TRAINING STATIONS

Air-to-Air Gunnery

Station	Training command instructors			Newly rated pilots			C. R.	Probability
	N	M	S. D.	N	M	S. D.		
Norfolk	40	8.4	3.9	156	7.6	3.3	1.10	0.14
Millville	42	3.9	2.1	126	3.6	2.3	0.66	.26
Suffolk	44	6.7	4.1	50	4.7	3.0	2.76	<.01
Combined probability (times occurring by chance in 1.00)								<.05

Air-to-Ground Gunnery

Station	N	M	S. D.	N	M	S. D.	C. R.	Probability
Norfolk	40	16.0	9.0	155	12.9	5.0	2.10	0.02
Millville	41	18.8	8.5	135	15.0	7.2	2.56	<.01
Suffolk	43	13.2	7.2	50	13.8	5.1	[1]-0.76	[1].33
Combined probability (times occurring by chance in 1.00)								<.01

[1] Note that these figures refer to a difference in the opposite direction.

Skeet Scores

1. *Description.*—A number of instructors in the Second Air Force expressed a belief that there was a high degree of correlation between skeet scores and scores in aerial gunnery. Since the former were relatively easily obtained at the stations involved, it was decided to make an exploratory study. In general each fighter pilot was required to fire at least 400 rounds of skeet during operation training. Scores were recorded in terms of percent hits obtained in sets of 25 rounds fired. At some stations scores of all sets of 25 rounds fired were recorded. At other stations records were maintained only for certain sets of 25 rounds spaced at about equal intervals throughout the 400 rounds fired.

2. *Distribution of scores and reliabilities.*—Data on skeet scores were

available for two classes at AAF Pocatello and one class at Peterson Field. These 107 fighter pilots had a mean score of 53 with a standard deviation of 9.6. The weighted average of the separate correlations between average odd and average even scores for the different classes was .33 giving an estimated reliability of .50 for the average total score.

3. *Evaluation.*—Although there was very little logical reason to suppose that proficiency in skeet shooting would be much related to over-all fighter pilot proficiency, the fact that this activity was included in the training requirements was felt to justify at least an exploratory study. For the small number of pilots studied, these scores are seen to have some low degree of reliability. The extent to which they are correlated with gunnery scores should give an indication whether or not the task is in any way related to gunnery performance. The average correlation between skeet and gunnery scores was found to be 0.28 for 96 pilots of three classes. This coefficient is rather high considering the unreliability of both types of scores. The small number of cases makes definite conclusions impossible.

Mission Grades and Comments

1. *Description.*—At some training stations in all three air forces responsible for training fighter pilots, it was customary to employ mission cards on which the instructors evaluated missions after their completion. On these cards the instructors usually assigned a grade and sometimes added comments on specific aspects of the student's performance. Grades were assigned for the most part according to the usual five point military rating scale ranging from "superior" to "unsatisfactory." At some station in the Fourth Air Force numerical percentage grades were assigned. The accompanying comments were sometimes favorable and sometimes unfavorable. The specific procedures used in grading, type of mission card employed and number of missions graded varied greatly from station to station and from one air force to another. In the First Air Force there were eight types of cards in use, one for each type of mission flown at the basic training stations (transition, formation, navigation, instrument flying, acrobatics, combat maneuvers, night flying and cross-country). In the Second and Fourth Air Forces a single general mission card was usually used. During the course of instruction, a student pilot was likely to be graded several times by each of several different instructors. In the First Air Force, grading was done by from 6 to 10 different instructors.

2. *Distribution of Different Mission Card Scores.*—Detailed studies of data obtained from mission cards were carried out only in the First and Fourth Air Forces. No mission grades or comments in the Second Air Force seemed sufficiently promising to warrant such analyses as were made in the other air forces. The frequencies and percents of different grades and categories of comments found in two independent studies of mission cards in the First Air Force are shown in table 2.17. In this particular analysis, neutral remarks were not counted.

In order to carry out further statistical analyses, research personnel in the First Air Force converted both grades and comments into quantitative scores. Weights for the various mission grades were obtained by assuming a normal distribution of pilot aptitude and then finding the distance from the mean in a normal curve for the actual percent receiving each rating.

TABLE 2.17.—*Distribution of instructors' grades and comments on individual mission cards*

FIRST AIR FORCE

First study (N = 200)

Grades	Frequency	Percent	Comments	Frequency	Percent
Superior	75	1.6	Commendations	3,036	69.6
Excellent	3,569	76.0	Neutral remarks	....	....
Very satisfactory	874	18.6	Minor criticisms	1,286	29.4
Satisfactory	145	3.1	Major criticisms	45	1.0
Unsatisfactory	31	.7			
Total	4,694	100	Total	4,367	100

Second study (N = 124)

Grades	Frequency	Percent	Comments	Frequency	Percent
Superior	76	1.7	Commendations	2,923	75.2
Excellent	3,773	86.2	Neutral remarks	....	....
Very satisfactory	423	9.7	Minor criticisms	928	23.9
Satisfactory	95	2.2	Major criticisms	34	.9
Unsatisfactory	12	.3			
Total	4,379	100	Total	3,885	100

These statistically derived weights were multiplied by a constant to remove decimal points and seven was added to each one to obviate working with negative scores. The final values employed were: superior = 11; excellent = 7; very satisfactory = 4; satisfactory = 2; unsatisfactory = 0. For the comments, favorable ones were given a value of 8; minor faults were given 4; and major faults, 0. Distribution statistics of the resulting grade-scores and comment-scores for three random samples of data are shown in table 2.18.

TABLE 2.18.—*Distribution statistics of proficiency scores derived from mission card grades and comments*

FIRST AIR FORCE

N	Grade Scores			Comment Scores		
	Range	M	S. D.	Range	M	S. D.
60	4.6–7.2	6.14	0.52	4.6–7.8	6.68	0.55
62	4.8–7.6	6.61	.48	5.2–8.0	7.00	.45
61	5.7–7.3	6.85	.30	4.4–8.0	6.73	.75

In table 2.19 are given the distribution statistics for scores derived from mission card grades and comments in the Fourth Air Force. The numerical values assigned grades and comments in the Fourth Air Force differed from those used in the First Air Force studies.

The grade scores yielding the data in the table were the raw numerical

grades given by the instructors. Mission comments were converted into numerical values by computing an "adversity score" for each pilot. This score was 100 times the ratio of the number of missions adversely commented upon to the total number of missions graded, one grade being given for each mission flown.

TABLE 2.19.—*Means and standard deviations of grade scores and comment scores from mission cards*

FOURTH AIR FORCE

N	Grade scores		N	Adversity comment scores	
	Mean	S. D.		Mean	S. D.
94	81.94	2.55	30	21.99	8.97
99	80.44	1.38	99	28.65	13.75

3. *Reliabilities.*—Grade scores and comment scores from odd and even numbered missions were correlated to obtain some estimate of the reliability of such scores. The reliability coefficients so obtained in the First and Fourth Air Forces are shown in table 2.20. The coefficients reported are surprisingly high in view of the following conditions: a. the total number of missions flown varied from student to student; b. missions were not flown in the same order by all students; c. the number of cards for any single type of mission was small; d. missions were flown with several different instructors; e. instructors varied widely in the number and type of comments recorded on the card; and f. instructors varied greatly in the relative degrees of importance they assigned to particular aspects of flying performance. It might be pointed out that the somewhat higher reliability

TABLE 2.20.—*Reliability coefficients of mission card grade scores and comment scores*

FIRST AIR FORCE

	N	Odd missions, 1		Even Missions, 2		r_{12}	r corr.
		Mean	S. D.	Mean	S. D.		
Grade scores:							
First study	105	6.07	0.85	6.10	0.72	0.68	.....
Second study[1]	162	2.40	.25	2.40	.26	.79	.....
Average r[2]	267	....	....	....	....	.75	.....
Estimated reliability	.....	....	....	....	....	......	.86
Comment scores:							
First study	104	6.67	0.90	6.83	0.70	0.56	0.72
Second study	150	6.94	.64	6.97	.66	.73	.84
Average r[2]	254	....	....	....	....	.67	.80

FOURTH AIR FORCE

Station							
Grade scores:							
VN-III	116	....	....	....	....	......	0.76
PV	33	2.85	0.15	2.86	0.15	......	.80
OV	117	....	....	....	....	......	.75
Weighted average	266	....	....	....	....	......	.76
Mission comments (no reliability data reported)	.....	....	....	....	....	......	.....

[1] Different weights used in two studies.

[2] By Fisher's z.

obtained for the mission grades as compared with comment scores might be a reflection of a "halo effect" in the assignment of grades.

4. *Interrelation of grades and comments.*—The relation between instructors' grades and their comments is shown in table 2.21 where the derived scores for grades and comments were correlated for several samples of data.

TABLE 2.21.—*Correlations between grade scores and comment scores from mission cards*

FIRST AIR FORCE

Sample	N	Grade scores, 1		Comment scores, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
SJI	28	5.60	0.50	6.72	0.81	0.60
SJII	50	6.16	.52	6.68	.55	.75
SJIII	62	6.61	.48	7.00	.55	.66
SJIV	18	6.25	.39	7.02	.41	.80
SJV	61	6.85	.30	6.73	.75	.58
Average[1]	229	......	......	......	......	.66

FOURTH AIR FORCE

Sample	N	Mean	S. D.	Mean	S. D.	r_{12}
FV	33	2.85	0.14	22.65	9.75	0.73
OV	117	80.71	1.51	29.60	14.65	.70
Weighted average	150	......	......	......	......	.71

[1] By Fisher's z.

5. *Evaluation.*—The First Air Force Unit made a comparison between experienced (former Training Command instructors) and inexperienced (recent graduates from the Training Command Flying Schools) pilots on scores derived from the mission card comments. The results appear in table 2.22. The difference in favor of the former instructors is evidence of the capacity of the mission card comments to differentiate experienced pilots from novices in the P-47 airplane.

The distributions of comment scores and the estimates of their reliability show that they should be adequate for use as criterion data of fighter-pilot

TABLE 2.22.—*Differences between former AAF Training Command instructors and newly-rated pilots in mission comment scores*

FIRST AIR FORCE

Comment scores[1]

	N	Mean	S. D.	C. R.	Probability
T. C. instructors	44	55	6.3	.....	....
Newly-rated pilots	200	50	9.8	4.0	<0.01

Grade scores[2]

	N	Mean	S. D.	C. R.	Probability
T. C. instructors	45	54	11	.....	....
Newly-rated pilots	164	49	13	2.4	<0.01

[1] T-scores were computed for all newly rated pilots at this station. The resulting conversion factors were used to derive the standard scores of the above samples of T.C. instructors and newly rated pilots of the same classes and commitments.

proficiency. While statistically the mission grade-scores appear to be as satisfactory criterion measures as those derived from the mission comments, the former are subject to the usual errors which attend general, over-all assessments of individual proficiency. The mission card *comments* have an advantage in that they are likely to be based on concrete instances of good and poor flying. Both grade and comment scores have the following advantages as criteria of proficiency:

a. They report the judgments and assessments of repeated flights, hence in the aggregate cover the entire course of training.

b. They represent a combination of relatively independent assessments of several instructors.

c. Being recorded immediately after the completion of the mission, the comments are relatively free from errors of memory.

6. *Development of experimental mission cards.*—Both the First and Fourth Air Force Units experimented with mission cards of their own construction designed for future introduction into all training stations as devices for obtaining proficiency data on fighter pilots. Research personnel in the First Air Force devised two tentative types of mission cards (*see* appendix B.1 and B.2) and tried them out at Richmond Army Air Base, then training commitment pilots who had just completed operational training in P-47 airplanes. The first card followed closely that already in use at one of the basic training stations in the First Air Force. Ten aspects of flying or gunnery were listed and there was left to the individual instructor the wording and details of the comments asked for in instances of particularly "good" or "weak" flying. The second card was of the check-list type and contained 19 items to be checked by the instructor where the student had manifested a weakness or an outstandingly good performance on that aspect of the flight. The cards were purposely made brief and undetailed since most of the instructors interviewed at several stations were agreed that a long, detailed card would be so time-consuming in use that it would lessen the effectiveness of the device for obtaining proficiency data. Experimental use of the two cards was curtailed by the sudden closing of the Richmond Army Air Base, followed shortly thereafter by the end of the war. Opinion of instructors was about evenly divided as to the relative value of the two cards. Statistics on use of each of the two cards for a brief period were as shown in table 2.23. The percentage of items checked and numbers of comments made are not in entire accord nor strictly comparable because some of the items on the cards apply only to particular types of missions. On the basis of the above data, it is not possible to determine which is the more suitable of the two types of card.

The Fourth Air Force Unit constructed a mission record form which was considerably more detailed than the cards described above. The check-list method was followed on this form. No data have been reported from actual use of this card at training stations.

Mobile Control Unit Records

1. *Description.*—At certain fighter pilot training stations of the First and Second Air Forces supervisory pilots in the control tower or in mobile control units placed near the runways, observed and recorded comments on the taxiing, take-off, flying in the traffic pattern, and landing of all aircraft operating at the station. Instances of gross errors, violations of safety regulations, and dangerous or careless handling of the airplane were supposed to be recorded. Since officers in the tower or mobile control unit did not know the identity of the pilots in the aircraft, items recorded should have been free from personal bias or "general impression" of pilot ability. In practice, there were extensive individual differences (between

Table 2.23.—Statistics on the use of two types of experimental mission card

FIRST AIR FORCE

Check-list type card	*N*	*Percent*
Number of cards used	172	...
Number of cards with items checked and/or comments	84	48.8
Total number of possible items	3096	...
Total number of items checked	274	8.9
Number of items checked as "weak"	52	1.7
Number of items checked as particularly "good"	222	7.2
Free response type card		
Number of cards used	122	...
Number of cards with comments	80	65.6
Total number of items for possible comment	1342	...
Total number of comments recorded	157	11.7
Number of adverse comments	32	2.4
Number of favorable comments	125	9.3

observers at a given station and from one station to another) in the fullness and accuracy of the records. This rendered very difficult any comparisons among different stations or different commitments of pilots. Since different officers were posted in the tower or control units from day to day, the resulting record of errors and violations, while indicative of common faults in basic flying technique, is probably somewhat unreliable as a source of criterion data.

2. *Distribution statistics.*—The number of errors recorded by mobile control unit operators was obtained for 9, 5 and 3 classes respectively at Millville AAF (1AF, basic training), Andrews Field (1AF, advanced training) and Strother Field (2AF, all training). Each class included from 25 to 80 pilots, the median size being 35. The average number of errors in landing recorded per class varied from 1.73 to 6.86 with standard deviations of from 1.56 to 5.17. For pilots in the 3 classes at Strother Field the percent of landings on which errors were noted ranged from 0 for some pilots to 80 percent for 1 pilot, the median being 30 percent. Six pilots

averaged more than one error for each landing made. Seventeen had no errors recorded. The variation from pilot to pilot and from class to class is striking. It seems doubtful whether the pilots in these classes would differ in average proficiency to the extent indicated by these figures. Probably many uncontrolled sources of error are present in such data.

3. *Reliabilities.*—The most adequate method of estimating reliability measures on the data described above would be to calculate the total number of errors for all the odd missions with those for all the even missions. This was not possible with the type of records available. For a sample of 100 pilots, research personnel in the First Air Force computed the correlation between the number of mobile control unit errors recorded during the basic and advanced phases of training. A product-moment r of 0.23 (uncorrected) was obtained. Individual differences in rate of improvement and station differences of all sorts would probably greatly reduce the size of coefficient to be expected.

TABLE 2.24.—*Differences between AAF Training Command instructors and newly-rated pilots in number of mobile control unit errors*

FIRST AIR FORCE

Commitment	T. C. instructors			Newly rated			C. R.	Probability
	N	Mean	S. D.	N	Mean	S. D.		
December	22	4.5[illegible]	[illegible]	29	4.69	3.59	2.35	<.01
January	24	2.23	[illegible]	26	4.41	5.27	4.16	<.01
February	29	[illegible]	[illegible]	59	3.77	2.92	.36	.36
March	15	2.65	[illegible]	44	3.15	2.48	1.33	.16
April	13	2.56	1.27	46	2.96	2.13	.54	.29

4. *Evaluation.*—A comparison of experienced (former Training Command Instructors) and inexperienced (recent graduates from Training Command Flying Schools) pilots, made in the First Air Force, yielded the results shown in table 2.24. The differences shown in the table are small for some commitments, but the inexperienced pilots consistently showed greater numbers of errors. Since the identity of the pilot against whom errors were recorded was not known by the personnel in the units, the above differences may be said to indicate some validity of the mobile control unit records as criterion data.

Perhaps the most serious criticism of reports of mobile control units as proficiency measures is the common belief that beyond a minimum level of proficiency, performance on landings and take-offs is relatively unimportant in combat flying. Although accepted by many, this belief is not supported by data on accidents which show that ability to land and take off safely is extremely important in combat training. In conclusion, it would be expected that if standard recording procedures were followed by all mobile control unit or tower operators and such were enforced at all training stations, the records of specific instances of faulty taxiing, take-off, traffic pattern flying, and landing would make useful proficiency data on

fighter pilots. However, this type of data would not be satisfactory if used as the sole criterion of fighter pilot proficiency.

Proficiency Ratings

1. *Printed rating scales. a. Description.*—After trial of a preliminary rating scale of five items at four training stations in the fall of 1944, revised rating scales for both the basic and advanced phases of training were employed in the First Air Force. Instructors rated each of the students on each item of the scale at the conclusion of training. Flight commanders were requested to confer with their instructors in arriving at a final proficiency rating for a student so that this rating would represent the pooled judgment of all instructors who had flown with him. Each item was rated on a five-point scale (5 = superior, 1 = minimum satisfactory). The items on the form used in basic training stations (*see* appendix B.3) fell into two categories, 18 items relating to *flying* technique, and 7 items relating to *personality traits* deemed important in fighter pilots. The rating scale used in advanced training stations (*see* appendix B.4) had three parts: 8 items on *flying proficiency;* 7 items on *gunnery technique;* and 7 items on *personality traits.*

b. *Distributions of ratings obtained.*—Ratings on the individual items of the scales were summed up to provide a total rating score; sub-total scores were also computed on each section of the rating scale (see above). The item counts and distribution statistics obtained indicated a slight tendency for the ratings to be concentrated at the high end of the scale and to cluster about the points 3 and 4. However, opinions of instructors, as well as the grades assigned on standardization board check flights at two stations, tended to corroborate the conclusions derived from the ratings—that relatively few pilots attract attention as being particularly "good" or "poor," and that there are more "good" than "poor" pilots in the usual run of commitments.

For two samples of 25 to 42 pilots at each of four Basic Training Stations the average total scores varied from 72.0 to 9[illegible] with standard deviations of from 7.4 to 16.6. Average scores for the flying items only, varied from 49.9 to 65.2 with standard deviations ranging from [illegible] to 11.6. Trait items showed considerably less variation from sample to sample with average scores between 22.1 and 28.2 and standard deviations of from 3.2 to 5.8. Similar data were available for two samples at each of four Advanced Training Stations. Total scores on the Advanced Scale averaged from 66.2 to 78.1 with standard deviations ranging from 6.2 to 1[illegible]8. Averages for flying items ranged from 24.2 to 27.9 with standard deviations of from 2.8 to 5.3. Gunnery items ranged from 20.6 to 24.8 in average scores and from 2.2 to 7.4 in standard deviation. Average scores on trait items ranged from 21.4 to 25.4 with standard deviations of from 2.5 to 4.6. Variations in score from one sample to another are seen to be considerably less with the Advanced Training ratings.

c. *Reliabilities.*—It was not possible to secure from the training stations more than a single set of ratings on individual student pilots. There remained two other methods of indirectly estimating the reliability of the ratings. The first of these involved intercorrelation of the part-scores on both the basic and advanced scales. The second approach was to correlate the total scores and part-scores on the basic scale with those on the advanced scale for the same sample of student pilots. The intercorrelation of part-scores for the two forms of the rating scale ranged from 0.60 to 0.92 for 3 or 4 samples from each of four Basic Stations. For two samples at each of four Advanced Stations the intercorrelations of part-scores ranged between 0.49 and 0.98 with a median of 0.66. The generally high intercorrelations for both scales are probably indicative of the "halo" effect

TABLE 2.25.—*Correlations between ratings assigned at basic and advanced training stations*

FIRST AIR FORCE

Millville May commitment (N = 40)

	Basic personality	Advanced flying	Advanced gunnery	Advanced personality
Basic flying	0.51	0.34	0.34	0.21
Basic personality		.30	.28	.22
Advanced flying			.66	.81
Advanced gunnery				.63

Norfolk April commitment (N = 30)

	Basic personality	Advanced flying	Advanced gunnery	Advanced personality
Basic flying	0.71	0.14	0.26	0.08
Basic personality		.18	.36	.16
Advanced flying			.63	.71
Advanced gunnery				.52

usually found in rating scales of this type, rather than evidence of reliability. As might have been expected, detailed examination of the ratings revealed differences in the adequacy of the ratings from one training flight to another, as indicated by range of ratings given, degree of independence of individual items, and correlation with other proficiency measures.

Correlations between scores derived from the basic scale and those derived from the advanced scale were computed for two commitments. Only nonexperienced pilots were included in the statistics, the results of which appear in table 2.25. The correlation coefficients obtained indicate some degree of relation between the ratings assigned at two different phases of training and by different groups of instructors.

c. *Evaluation.*—As with other criteria, a comparison was made between experienced and inexperienced pilots for several samples of students on both the basic and advanced forms of the scale. Results are shown in table 2.26.

It is difficult to tell whether the striking differences in the table favoring the former instructors were indicative of the adequacy of these ratings as proficiency data or were merely an indication of a generalized halo effect

operating in favor of the more experienced pilots. Any tendency to base ratings upon experience rather than ability would also operate to raise the correlations obtained between ratings on separate parts of the scales and between ratings at basic and advanced stations. Follow-up visits to the training stations where the scales were in use indicated that several of

TABLE 2.26.—*Differences in proficiency ratings (basic and advanced scales) for AAF Training Command instructors and newly rated pilots*

FIRST AIR FORCE

RATING SCALE USED AT BASIC STATIONS

Station	T. C. instructors			Newly-rated			C. R.	Probability
	N	M	S. D.	N	M	S. D.		
	Flying proficiency items							
Andrews	39	58.8	6.9	117	53.7	7.4	3.92	<0.01
Bradley	37	65.8	12.2	91	58.0	8.4	3.57	< .01
Richmond	[illegible]	70.6	9.7	86	63.6	8.5	3.26	< .01
Seymour Johnson	41	60.0	8.8	152	57.4	9.1	1.62	.06
	Trait items							
Andrews	39	27.1	4.4	117	23.5	4.4	4.41	<0.01
Bradley	37	27.0	4.8	91	24.5	4.7	2.73	< .01
Richmond	25	28.7	4.4	86	25.7	4.3	3.02	< .01
Seymour Johnson	41	26.2	4.5	152	24.9	4.5	1.60	.06

Combined probability (times occurring by chance in 1.00)—less than .01.

RATING SCALE USED AT ADVANCED STATIONS

Station	T. C. instructors			Newly-rated			C. R.	Probability
	N	M	S. D.	N	M	S. D.		
	Flying Proficiency Items							
Dover	44	31.3	4.9	96	27.3	4.6	4.56	<0.01
Millville	101	27.3	4.8	[illegible]	24.9	4.0	4.31	[illegible].01
Norfolk	65	28.7	4.3	[illegible]	26.9	3.9	2.71	[illegible]
Suffolk	58	28.3	5.4	56	25.8	3.5	2.98	< [illegible]
	Gunnery Proficiency Items							
Dover	44	25.1	3.2	96	23.9	[illegible]	2.10	0.04
Millville	101	23.2	3.8	194	21.3	[illegible]	4.22	< .01
Norfolk	65	24.7	3.7	117	23.0	3.5	2.97	< [illegible]
Suffolk	58	22.8	3.8	56	22.0	3.5	[illegible]	.25
	Trait Items							
Dover	44	27.4	4.4	96	25.4	4.1	2.52	<0.01
Millville	101	21.4	4.1	194	22.5	4.0	3.89	[illegible].01
Norfolk	65	25.6	4.0	117	24.1	4.7	2.31	.02
Suffolk	58	25.8	4.8	56	24.5	4.3	1.56	12

Combined probability (times occurring by chance in 1.00)—less than .01.

the more common sources of error, which usually operate where rating scales are used, were also influencing these ratings. The most serious drawback was the lack of specific cause on which to base the assigned ratings. The doubtful value of the printed proficiency rating scales led to the decision to substitute for them a system of interviewing instructors at the training stations.

2. *Instructor-interview ratings.* a. *Description.*—For four commitments of student pilots, aviation psychologists in the First Air Force visited the advanced training stations and secured ratings of proficiency through de-

tailed interviews with the instructors at the conclusion of training. Four types of information were secured from each instructor on the student pilots in his flight: selection of the most proficient and the least proficient pilot in the flight; assignment of an over-all proficiency rating of pilot ability on a five-point scale; the approximate number of missions flown with each student; the specific reasons behind the assigned ratings in each case. In giving reasons for the ratings instructors were required to cover all of the major aspects of flying training (general handling of the airplane, formation flying, execution of maneuvers difficult to follow, aerial gunnery technique, ground gunnery and strafing technique, freedom from accidents

TABLE 2.27.—*Reliability coefficients of proficiency ratings obtained from interviews with instructors at advanced training stations*

FIRST AIR FORCE

Commitment and station	N	Mean	S. D.	r_{tt}
May:				
Dover	56	3.33	1.38	0.38
Millville	60	3.32	1.72	.62
Norfolk	65	3.29	1.59	.49
Suffolk	20	3.65	1.44	.78
June:				
Bluethenthal	50	3.77	1.56	.51
Dover	55	3.60	1.40	.72
Millville	43	3.51	1.39	.76
July:				
Bluethenthal	61	3.23	1.26	.76
Dover	51	2.54	1.21	.49
Millville	42	3.48	1.49	.61
August:				
Bluethenthal	46	3.86	1.25	[1].80
Dover	43	3.88	1.30	.65
Millville	[2]39	3.52	1.45	.76
Average (by Fisher's z)	631	.....	.....	.65

[1] Correlation between ratings secured by two different interviewers.
[2] Inexperienced pilots only.

or near-accidents, attitude and air discipline). Particular emphasis was placed on reasons given for extremely high or low proficiency ratings. These comments or explanations of the ratings were entered in the notes of the interview.

b. *Distributions and reliability of interview ratings.*—The ratings obtained in the interviews were given numerical values as follows: "Above average," 5; "Slightly above average," 4; "Average," 3; "Slightly below average," 2; "Below average," 1. In addition, if a pilot was selected as the best pilot in the flight, the numerical rating given was increased by 1. If selected as the least proficient man in a flight, one point was subtracted from the rating. The ratings of the several instructors on a single student were averaged to yield a single proficiency measure. Means and standard deviations of ratings, together with reliability coefficients obtained by correlating the ratings secured from two different instructors are given in table 2.27 for the proficiency ratings on students in four commitments.

c. *Evaluation.*—Comparison of experienced and newly rated pilots on

the instructor-interview ratings for student pilots in two commitments yielded the results in table 2.28.

The advantages of the instructor-interview ratings over those obtained by use of the printed scales are the following:

(1) Specific reasons can be found for the assignment of ratings. This is especially important in cases in which ratings are obviously influenced by personality factors rather than by pure flying ability.

(2) Degree of acquaintance with the students' flying can be taken into account; instructors are excused from rating any student in the flight with whose flying they are not sufficiently familiar.

(3) The interviewer has an opportunity to note the relative ability and willingness of the several raters to estimate the proficiency of the students.

TABLE 2.28.—*Differences in proficiency ratings (obtained from instructor interviews) for AAF Training Command instructors and newly-rated pilots in advanced training stations*

FIRST AIR FORCE

Station	Experienced			Newly rated			C. R.	Probability[1]
	N	Mean	S. D.	N	Mean	S. D.		
May commitment								
Dover	9	3.64	0.79	47	3.53	1.00	0.74	0.23
Millville	9	4.08	1.60	38	3.41	1.36	1.26	.11
Norfolk	14	3.96	1.37	41	3.56	.84	1.26	.11
Suffolk	5	4.85	.96	20	3.73	1.04	2.10	.02
June commitment								
Dover	12	4.63	.89	35	3.67	.97	2.94	<.01
Millville	15	[illegible].28	1.25	35	3.43	1.20	2.20	.01
Norfolk	10	[illegible].30	1.34	40	3.75	1.10	1.01	16

[1] Combined probability (number of chances occurring in 1.00) < .01.

The ratings of different instructors can then be weighted accordingly in arriving at the total rating for the student.

In the data here reported the third advantage given above was not utilized. All instructors who were able to rate a student did so and their ratings were given equal weight in determining the final result. The extent to which instructors changed their ratings because of having to justify the ratings by specific reasons is not known. Interviewers reported such changes were fairly frequent.

3. *Rankings by instructors.*—The instructors at three stations of the Second Air Force were asked to rank their fighter pilot students in the order of their expected over-all effectiveness in combat. The rankings were then converted to percent-position scores for validation studies. Since independent ratings of the same students by more than one instructor were not available, no reliabilities were computed for this criterion. Because instructors were asked to make an over-all judgment and no guidance was given as to specific items to be considered, it is probable that the rankings were based to a great extent upon subjective factors. Also the basis of judgment undoubtedly varied greatly from one instructor to another.

Such rankings do have the advantage of obtaining a single over-all judgment of proficiency.

4. ***Ratings by fellow students.*** *a. Description.*—On a sample of pilots who had completed operational training, research personnel in the Second Air Force secured proficiency ratings from their fellow students. Each pilot was asked to list those of his classmates he felt he knew well enough to rank in order of proficiency as combat pilots. He was told to list first the pilot he thought would be most proficient, and to place the pilot he thought would be least proficient last, with the rest in order in between. With this procedure, pilots differed greatly in the number they listed. Hence, there was also great variation in the number of ratings different pilots received. What sort of bias may have influenced the individual pilots in their choice of pilots to be listed was not known. Statistical treatment

TABLE 2.29.—*Correlations between percent-position scores derived from ranks assigned by odd and even raters*

FOUR OR MORE RATINGS—SECOND AIR FORCE

Station	Number cases	Percent position—odd raters, 1		Percent position—even raters, 2		r_{12}
		Mean	S. D.	Mean	S. D	
Peterson	36	51	1.2	51	1.2	0.54
Abilene	30	49	1.5	48	1.3	.79
Ft. Sumner	8	50	1.3	44	1.3	.36
Harding	19	47	1.2	53	1.2	.53
Strother	35	49	1.1	50	1.1	.78
All stations (average by Fisher's z)						.65

of the data was restricted to the rankings of those pilots who had been listed by four or more of their classmates.

b. *Reliabilities.*—Correlations were computed between the average percent-position scores for the listings of odd and even classmates. The coefficients are listed in table 2.29, for those stations from which data were obtained. The percent-position score in each case was the percentile rank corresponding to the position of the pilot in the list of one of the raters. Fifty-three of the student pilots were rated by eight or more classmates. The odd-even reliability for this group of ratings was .74. These data show that ratings by a number of fellow students may be expected to yield estimates of proficiency of sufficient reliability to the be useful measures of proficiency. Whether the bases of rating were valid—i.e., were measures of combat proficiency—could not be determined with the data at hand.

Flying Evaluation Board Reports

Analysis of records on fighter pilots who were called before Flying Evaluation Boards have been reported from the First, Second, and Fourth Air Forces. Results of the analysis of Flying Evaluation Board cases were reported in terms of the cause for meeting the board (e.g., lack of proficiency, fear of flying, physical defect, disciplinary reasons, etc.). Since data

from these studies have already been presented in another section in the discussion of job specifications for fighter pilots (see page 25) and since there was no method of determining their reliability, no further presentation of these data is necessary here. The small number of cases re-evaluated in all of the four continental air forces places severe limitations on the usefulness of this criterion of fighter pilot proficiency.

Aircraft Accidents

1. *Description.*—Surveys of aircraft accident reports have been made by all four of the Air Force units. The approach in all cases was to analyze the information from the Accident Board Reports on file at the Air Force Headquarters and then to set up various accident categories for subsequent comparison as to stanines and classification test scores. The categories selected for stanine validation varied somewhat from one Air Force unit to another, but generally included several of the following:

a. Accident cases of all kinds
b. Pilot-error accident cases
c. Non-pilot-error accident cases
d. Accident cases of unknown causation
e. Accident-free population (or an accident-free control group)
f. Single accident cases and multiple accident cases
g. Fatal accident cases and non-fatal accident cases
h. Non-injury or slight-injury-to-personnel cases
i. Cases of severe injury to personnel
j. Cases invoiving slight damage to aircraft
k. Cases involving heavy or complete damage to aircraft

Since the several categories were not exactly comparable from one air force to another, a summary table of frequencies in each category could not be derived from the data which were available.

2. *Reliability.*—Because of the small number of pilots who were involved in more than one accident due to pilot error, there is no feasible method of estimating the reliability of aircraft accidents as a criterion of fighter pilot proficiency.

3. *Analysis of psychological reasons for accidents.*—A detailed psychological analysis of pilot-error accidents was attempted by First and Second Air Force research personnel. The results of their analyses have been presented in an earlier section of this chapter (see page 25) in connection with the discussion of job specifications. One serious difficulty in the way of arriving at *significant* categories of accident causation lay in the nature of the Accident Board Reports, in which the degree of pilot responsibility was frequently difficult to assess from the information given. Though the separate agreement among the psychologists in the two Air Forces, who acted as a jury in making the analysis of the accident records, was fairly satisfactory in both, the value of the resulting trait frequencies is perhaps open to debate. Such agreement could be brought about by stereotyped ideas of

the nature of certain accidents operating either in the psychologists acting as the jury or in those reporting the details of the accidents. Besides, the small number of available cases restricts the value of any conclusions suggested by these analyses. Certainly there were extensive disagreements between the distributions obtained in the two Air Forces.

In the Second Air Force a special study was made of the degree of agreement between different judges in the psychological reasons ascribed to the accidents. In table 2.30 are listed the frequencies of agreement and disagreement among four psychologists using successively 5, 8 and 15 categories of reasons. There was obviously greater agreement when a few general psychological categories were used than when there was a larger number of more specific categories from which to choose. Since one would expect a greater degree of agreement just by chance when categories are few, coefficients of contingency were computed to express the extent to which the observed agreement differed from chance expectations. Separate

TABLE 2.30.—*Percent of accidents classified in the same manner by four judges*

P-47 ACCIDENTS—SECOND AIR FORCE

Degree of agreement	Number of categories used		
	5	8	15
All four judges agreed	47	36	26
Three judges agreed	33	32	33
Two judges agreed	20	32	35
No agreement	0	0	6
Total	100	100	100

coefficients were computed for the comparisons of the classifications of each pair of judges. The average coefficients obtained for all pairs were 0.56, 0.71 and 0.76 for the 5, 8 and 15 category classifications respectively. That the greatest superiority over chance agreement occurred with the 15 category classifications was contrary to expectations from the data in table 2.30. Even discounting the fact that the maximum possible coefficient is higher for comparisons based on larger numbers of categories, there still remains evidence of better agreement in discrimination of reasons for accidents when the categories are many and specific. The data are not sufficient to provide any further evidence on this point.

4. *Evaluation.*—It is difficult to evaluate aircraft accidents as a criterion of fighter pilot proficiency. Opinions of instructors and of flying safety officers varied widely as to whether or not accident occurrence was indicative of low pilot aptitude. It was commonly asserted by supervisory personnel in the training stations that "any pilot can have an accident!" On the other hand, some felt that almost all accidents, even those reported as caused by materiel failure, were evidence of some lack of proficiency on the part of the pilot. However, it is logical to assume that large numbers of accidents in operational training originated from circumstances over which the pilot had no control, such as materiel failure, adverse weather conditions,

etc. Even in accident cases where pilot error may have contributed to the severity of the accidents, there is considerable disagreement by the pilots on the accident boards regarding the assessment of responsibility.

The number of pilot-error accidents was too small to warrant very much emphasis upon a statistical study relating accident occurrence to other measures of pilot proficiency such as gunnery scores, mission card comments, proficiency ratings, tower reports, etc. Furthermore, the number of accident repeaters is so small as to preclude any decision as to the reliability of accident occurrence. The only statistical evidence regarding the usefulness of this criterion is given in a First Air Force study comparing the accident frequency among experienced and inexperienced pilots. These data are given in table 2.31.

TABLE 2.31.—*Pilot-error accident occurrence among experienced and newly-rated pilots*[1]

FIRST AIR FORCE

	Total N	P-E accident cases	Per-cent	C. R.	Prob-ability
Experienced pilots (former training command instructors)	371	23	6.2	......	......
Inexperienced pilots (recent graduates from Training Command flying schools)	1110	110	10.0	2.53	<0.01

[1] Student pilots in commitments from October 1944 through July 1945.

Time Required to Complete Training

1. *Description*.—There are extensive individual differences among student pilots in the time elapsing between the start of operational training and the completion of training and assignment to staging stations. Time may be lost in the basic and advanced phases of training because of many factors: grounding because of physical disability, prolonged adverse weather conditions, inadequate number of operational aircraft, emergency leaves, changes in commitment requirements, etc. Except for the first and last of these factors, however, their influence upon a population of student pilots is probably random. The belief is often expressed by supervisory personnel in field training stations that more proficient pilots complete their training (on the average) more rapidly than less proficient pilots. Studies were undertaken in the First and Second Air Forces to discover the relation of pilot stanine to the time required to complete operational training in the P-47 airplane.

In the First Air Force study three measures of time-in-training were used: (1) number of months required to accomplish basic training; (2) number of months spent in advanced training; (3) number of months required to complete basic and advanced training combined. Student pilots with no previous experience as instructors who completed training in the May, June, and July commitments in the I Fighter Command were used in the study. In the Second Air Force three scores were derived from the training progress charts for pilots in the December 1944 commitment at the end of their fifth week of training. One score represented a weighted

score based on number of missions reported (repeated missions were given less weight), another was the number of hours flown and the third represented the number of AAF standards completed. The data from both air forces revealed that in all of the scores obtained there was considerable variation between stations and between classes or commitments at the same station.

2. *Reliabilities.*—There seemed to be no feasible method of estimating the reliability of the above measures other than by correlating the time spent in basic training with that spent in advanced. In three commitments of more than 100 pilots each the coefficients ranged from —0.42 to 0.50 with an average of —0.02 for the 356 pilots in the total group. The data showed a zero relationship between time spent in the two phases of training. Investigation of samples of data from different training stations revealed that there were unexplained but fairly consistent (and often

TABLE 2.32.—*Differences in time required to complete operational training for Training Command instructors and newly-rated pilots*

FIRST AIR FORCE

Training phase	T. C. Instructors			Newly rated			C. R.	Probability
	N	Mean	S. D.	N	Mean	S. D.		
Basic only	44	4.07	0.77	201	4.46	0.81	3.00	<0.01
Advanced only	39	2.15	.36	144	2.19	.39	.60	.27
Basic plus advanced	44	6.32	.94	201	6.83	.91	3.29	<.01

statistically significant) differences from station to station. These differences probably operated to obscure the true significance (if any) of time-in-training as a possible criterion of fighter pilot proficiency. Time spent in basic training *did discriminate* between experienced and inexperienced pilots, as shown in table 2.32. The data for advanced training are less clear cut.

4. *Evaluation.*—In evaluating time-in-training as a possible criterion, the following points have to be considered. First, it was extremely difficult to obtain records of actual flyable time as distinguished from mere presence at a training station. Administrative reasons for loss of flying time and legitimate grounds for periods of indefinite suspension from flying further complicated the picture. It would be necessary to make a detailed examination of each student pilot's Form 5 and to secure the needed information from the supervisors of flying training (relative to lack of aircraft, poor weather, commitment requirements, etc.) to obtain a thorough and accurate analysis of the time-in-training criterion.

Furthermore, in connection with use of this criterion in validity studies, it has been shown that for each successive class graduating from the Training Command since Class 43-G, the average stanines have been higher. (*See* report number 2). Thus, even though there were no genuine relationship between time-in-training and pilot stanine, a spurious relationship might be obtained from the data, since in the same I Fighter Command commitment,

men from the earlier Training Command classes would show, on the average, both a longer time-in-training and a lower stanine. Against this interpretation is the fact that the experienced and inexperienced pilots showed a clear-cut difference in time taken to complete the basic phase of training. Final evaluation of time-to-complete training and similar measures would require additional study and analyses designed to control the effects of some of the important factors influencing such a measure.

Retention of Students as instructors

During the latter part of 1943 and continuing up to about August 1944, a number of student pilots in the I Fighter Command were retained at the basic and advanced training stations after the completion of operational training as instructors to replace those moving on to combat theaters. Supervisory personnel at the training stations agreed that these trainee pilots were selected as instructors for having demonstrated unusual proficiency during their period of training. A total population of 209 of these pilots was uncovered after search through the files at Headquarters, I Fighter Command. This population was taken as a selected group of high-proficiency student pilots against which to validate the stanine.

As a criterion of proficiency, retention as an instructor is open to a number of objections. There was almost no way of determining the reliability of the decisions made. Also, the numbers involved were so small that no extensive studies could be made. It is likely that some subjective factors entered into the decision, and the basis of decision may have varied from one station to another. However, as a group, the students retained as instructors were undoubtedly outstanding in proficiency. Finally, the over-all type of judgment required in deciding which pilots were retained was probably based upon excellent coverage of all aspects of the job.

Written Proficiency Tests

1. *Description.*—At the request of the Wing Director of Training, research personnel from the Second Air Force aided in the development of a written proficiency test for P–47 pilots. After construction of the test by Instructors of the Wing, the research personnel did a statistical analysis of the results of a preliminary administration to 48 fighter pilots who had completed operational training. The test consisted of 373 items distributed among eight parts or sections.

2. *Reliability.*—The odd-even reliability coefficients for each of the eight parts of the test were as given in table 2.33 below.

An item analysis was also carried out, from the results of which recommendations for revision of the test were made to the Wing Director of Training.

3. *Evaluation.*—Although written proficiency tests meet most of the requirements of a good criterion (such as reliability of scores, objectivity, and high degree of discrimination between individuals), there is one im-

portant requirement in which they are deficient. They do not adequately cover the job of the fighter pilot; and there is little evidence as to how closely scores on such tests are related to success in combat operations. For these reasons scores on written proficiency tests can only be used to supplement other measures of proficiency. Used in this manner, they should have some value in measuring fighter pilot proficiency.

Combinations of Criteria

Where no one criterion provides an adequate measure of proficiency it is sometimes possible to meet the requirement of adequate coverage of the job by combining various criteria, each of which is based upon only one of several of its important features. Thus several combinations of criteria were tried out in the various Continental Air Forces. The most extensive

TABLE 2.33.—*Odd-even reliability coefficients for written proficiency test*

48 P-47 PILOTS—SECOND AIR FORCE

Part	r_{tt} corrected
A. Nature and characteristics of the aircraft	0.43
B. Orientation and intelligence	.12
C. Regulations and technical subjects	.56
D. Recognition (air and sea)	.33
E. Armament and gunnery	.15
F. Tactics	.67
G. Survival	.28
H. Military subjects	.51
Total test	.79

combinations were used in the First Air Force. In a preliminary study the separate criterion scores were in each case combined after transmuting raw scores to standard scores with a mean of 50 and a standard deviation of 10. Combinations were first made of air-to-air and air-to-ground gunnery scores to form a single gunnery criterion for all pilots on whom both measures were available. For the May, June, July and August 1945 commitments, combinations were made of gunnery scores, ratings from basic stations and interview ratings from advanced stations. Finally, a combination of all available criteria was used to obtain two extreme groups of "high" and "low" proficiency for use in test validation. Details of the various procedures are given in the section on Validation. In the Second and Fourth Air Forces a two-category criterion was obtained by combining pilots with aircraft accidents and re-evaluated pilots into failure groups for use with groups of successful pilots as a pass-fail criterion in operational training. Since the number that had accidents and the number re-evaluated were both small, such a procedure had the advantage of providing a larger number in the failure category for any two-category statistical computations.

General Evaluation

Relations Between Criteria

Data on intercorrelations among the several proficiency measures on fighter pilots have been reported from the First, Second, and Fourth Air Forces. These statistics have an important bearing on the decision as to the identity and number of separate criteria needed to serve adequately

TABLE 2.34.—*Intercorrelations among various criteria for P-38, P-47 and P-61 pilots*[1]

P-38 PILOTS—FOURTH AIR FORCE

Criterion	1	2	3	4	5	6	7[2]
Air-to-air gunnery....1		0.11	0.06	0.[illegible]	0.[illegible]	0.06	−0.[illegible]
Air-to-ground gunnery....2	938		−.03	.11	.0[illegible]	.0[illegible]	.1[illegible]
Sk[illegible]ing....3	253	240		.21	.25	.0[illegible]	−.1[illegible]
[illegible] bombing....4	660	375	234		.12	−.12	.15
[illegible]se grades....5	225	19[illegible]	43	143		.71	.1[illegible]
[illegible]versity ratio....6	129	129	32	32	150		.[illegible]0
Accidents....7	850	677	204	333	236	106	

P-47 PILOTS—FIRST AIR FORCE

Criterion	1	2	3	4	5
Air-to-air gunnery....1		0.21	....	....	....
Air-to-ground gunnery....2	436		....	....	....
Combined gunnery....3	....	....		.16	.29
Basic rating....4	....	....	313		.21
Advance interview rating....5	....	....	312	313	

P-47 PILOTS—SECOND AIR FORCE

Criterion	1	2	3	4	5	6	7
Air-to-air gunnery....1		0.40	0.37	.....	0.28	0.17	.....
Air-to-air gunnery....2	110		−.04	0.45	.....	.....	.....
Dive bombing....3	79	73		.....	.....	.....	.....
Skip bombing....4	....	73	79		.....	.....	.....
Sheet scores....5	96	.....	.....	.....		.....	.....
Instructors rankings....6	72	.....	.....	.....	.....		0.54
Mutual rankings....7	.....	.....	.....	.....	.....	51	

P-61 PILOTS—FOURTH AIR FORCE

Criterion	1	2	3	4	5	6	7
Air-to-air gunnery....1		0.24	−0.13	.....	.....	.....	.....
Air-to-ground gunnery....2	122		−.04	.....	.....	.....	.....
Aircraft accidents....3	184	202		.....	.....	.....	.....

[1] Entries below the diagonal are the number of cases for all corresponding correlation coefficients.
[2] Biserial coefficients of correlation.

for classification test validation purposes. The results of intercorrelation studies reported from the various Air Forces are summarized in table 2.34. Except where otherwise indicated the coefficients listed are averages (by Fisher's z) of separate coefficients for the different classes and stations.

The correlation coefficients obtained among the various proficiency measures on P-47 pilots, while generally low, are consistently positive from sample to sample and appear to be indicative of a genuine relation among the measures. It is difficult to determine to what extent the low magnitude of these correlations reflects true independence of the various

criterion measures and how much it results merely from the low reliability of the measures themselves. If these results be taken to indicate considerable independence of the several criteria, it must be concluded that no one or two of them adequately covers the area of fighter pilot proficiency. The data reported from the Fourth Air Force must be interpreted in the light of the low reliability (0.31) of the air-to-air gunnery scores obtained there. What evidence there is would indicate that mission grades and various types of ratings have the most in common with the other criteria.

Over-all Evaluation

1. The quantitative scores derived from gunnery, bombing and rocket firing are probably the best criteria under the standards on page 24. Their distribution is fair and has improved with recent emphasis on gunnery training methods and accuracy of recording; the reliability, while not high, is acceptable; they discriminate between experienced and inexperienced student pilots; while subject to a number of clerical errors, they are free from the more important errors of subjective judgment; they cover an important part (though not all) of the area of fighter pilot training.

2. The data from individual mission cards provide another good criterion source. This is especially true of the comments made by instructors on specific behavior observed during the flight. Pilot proficiency scores derived from the comments, though poorly distributed, have adequate reliability, discriminate between experienced and inexperienced pilots, are relatively free from subjective judgment errors (being related to actually observed instances of flying) and cover all aspects of training since they can be made after every flight. Grades on the mission cards provide less satisfactory criterion data, being too subject to subjective impressions, "halo" effect, and hasty generalizations; their distribution is very unsatisfactory, there being far too little scatter of assigned grades.

3. The mobile control unit (and control tower) reports afford criterion data which potentially might serve to supplement other criterion measures. As received from the training stations, distribution is not satisfactory; there is no measure of reliability; they do discriminate successfully between experienced and inexperienced pilots; they are free from errors of personal bias, but are subject to careless recording; they sample only a relatively small and less important part of flying in fighter type operational aircraft.

4. Proficiency ratings assigned at the end of training on printed forms may be rejected as being too susceptible to all sorts of subjective errors, failing to cover the whole course of training adequately, and being poorly distributed. Proficiency ratings secured from interviews with instructors at the completion of training have several advantages over those assigned on printed forms. As secured in the First Air Force, the distribution of ratings is good; reliability (agreement among instructors) is satisfactory; they discriminate between experienced and inexperienced pilots; they obviate

subjective errors to a great degree by being supported by specific evidence; they cover the course of training well since several instructors are interviewed to include all types of missions. These ratings are no doubt somewhat subject to hasty generalization, and also to memory errors. However, they probably pick out the best and poorest pilots quite accurately, since these students have made more impression upon instructors.

5. Insufficient research has been done on aircraft accidents to enable an assessment to be made as to the adequacy of this criterion. The distribution is poor; no estimate of reliability is possible; adequate data are not available as to the comparative accident frequency among experienced and nonexperienced pilots; assessment of pilot responsibility by the Accident Boards is not entirely free from errors of subjective judgment; accidents may (and do) occur at all stages of training and experience.

6. Cases of pilots called before the Flying Evaluation Boards for flying deficiency might be expected by many to provide an excellent criterion, comparable at the operational level of training to the graduation-elimination criterion in the Training Command flying schools. Practically, the number of cases re-evaluated in the Continental Air Forces is too small to furnish an adequate criterion for classification test validation. The small number of men re-evaluated for flying deficiency in operational training probably reflects in part training policies and partly the previous screening at the classification centers and flying schools of the Training Command. This criterion could probably be greatly improved if there were some means of identifying border-line cases, both those that were eventually reclassified and those that were returned to flying status. Also some means of estimating the degree of reliability of FEB action would permit a better assessment than is now possible of its usefulness in criterion studies.

7. It is difficult to evaluate time taken to complete training as a criterion since there is no practical way of estimating its reliability. It is subject to a host of variable factors which would prove most difficult to control. It has the advantage of being objective and relatively free from subjective errors. It also provided an overall assessment based on most aspects of operational training. This criterion probably more nearly achieves an objective measurement of the important factor of motivation than does any other measure so far studied.

8. The foregoing survey of proficiency measures would seem to warrant the conclusion that among records currently available gunnery and bombing scores, and mission card comments provided the best criterion measures of fighter pilot success in operational training. Proficiency ratings secured by interview with instructors also appear to have afforded a good criterion. It further seems apparent that no one of the above criteria covered all aspects of fighter pilot training comprehensively enough to serve as the sole instrument for validating the classification tests. From a superficial evaluation of the available data it would seem that the smallest number of separate proficiency measures which would accomplish this purpose would be

the objective records obtained from gunnery, bombing and rocket firing and the specific comments recorded on mission cards. It might be pointed out that were uniform regulations for the recording and collection of these proficiency data to be issued and enforced in all commands the statistical handling of such data would be somewhat simplified and greater reliability of scores should result from the elimination of some of the now uncontrolled variability.

VALIDATION

One of the primary goals of the Aviation Psychology Program in the Continental Air Forces was to validate the procedures used in the AAF Training Command in the selection and classification of aircrew officers. However, because of difficulties encountered in obtaining adequate measures of proficiency, research personnel working with fighter pilots and bombardment crews were forced to spend considerable time and effort in searching for and developing criteria of proficiency. The results of such activity with fighter pilots have been discussed in the previous section on Criteria. Only after these studies had been accomplished was it possible to determine adequately the relationship between the aptitude scores or stanines of the fighter pilots and their degree of success in operational training. It had also been intended that studies would be made of the relation between scores on the individual classification tests and success in operational training. However, only a few studies of the validity of individual classification tests were accomplished.

In the discussions of validities of stanines and tests in this and succeeding chapters all validity coefficients that are significant at the 1 percent level (i.e., could occur by chance less than 1 time in 100) are identified with two asterisks. Those coefficients that are significant at the 5 percent level are marked with a single asterisk. Validities of stanines are presented first, followed by data on test validities at the end of the section. Separate coefficients are presented for each of the criteria studied, as well as for certain combinations of criteria.

Stanine Validation

Fixed Gunnery, Bombing and Rocket Firing

The ability to hit a moving or stationary target with either guns, bombs or rockets is the ultimate aim of all fighter pilot training. Fixed gunnery, bombing and rocket firing scores thus naturally become the most obvious measures of fighter pilot proficiency. Most of the validity studies of the pilot stanine have therefore been made against these criteria.

The correlations of the pilot stanine against aerial and ground gunnery scores for all stations and commitments combined are shown in table 2.35 for pilots in the First, Second and Fourth Air Forces.

These coefficients were computed separately for each class or commitment and station and were averaged by Fisher's z to produce the coefficients

been reported. Examination of the separate coefficients for different classes in the First Air Force revealed that slightly more than half of them fell within plus or minus 1 P.E. of 0.15, assuming an N of 50 (approximate average size of class). Distribution of validity coefficients in the other Air Forces appeared to be similar in this respect. The validity of pilot stanine against gunnery T-scores in the First Air Force which represented an equally weighted combination of ground and air-to-air gunnery scores, was found to be .14** or slightly less than the validity against air-to-air gunnery alone.

TABLE 2.35.—Correlations of pilot stanine and aerial and ground gunnery scores

Air force	Type plane	Aerial gunnery N	Aerial gunnery r	Ground gunnery N	Ground gunnery r
First	P-47	1,270	0.13**	1,270	0.10**
Second	P-47	407	.16**	66	.13
Fourth	P-38	949	.11**	609	.14**
Fourth	P-61	204	.19**	346	.04

It has been demonstrated in the section describing the criteria that gunnery scores tend to fall toward the low end of the scale of scores. Also, because of the selective processes operating throughout the student officers' period of training, many men with low stanines have been eliminated before reaching operational training, so that the stanine distribution of all men in training in the Air Forces is skewed in the opposite direction from the gunnery scores, most of the pilots having high stanines. A situation exists, therefore, in which the chances are highly favorable for high stanine individuals to obtain low gunnery scores. This situation, plus the unreliability of the criteria, partially explains the low correlations obtained between the stanines and gunnery scores. A truer picture of the predictive ability of the stanines may be seen from an examination of table 2.36 which shows the distribution by stanine of combined T-scores representing aerial and ground gunnery for newly rated pilots only. It may be seen from this table that some high stanine men make just as low scores as low stanine men, but low stanine men do not make as high scores as men with

TABLE 2.36.—Distribution by stanine of combined T-scores representing aerial and ground gunnery

P-47 PILOTS—FIRST AIR FORCE

Newly-rated pilots only

Stanine	N	Range	Mean T-score	Percent exceeding mean T-score of total group
9	115	75—153	107.90	[illegible]
8	184	69—159	102.34	47.28
7	225	69—141	102.01	47.11
6	287	75—147	99.86	41.11
5	233	69—160	97.96	38.19
4	75	69—123	92.72	24.00
3	31	81—129	93.68	19.35
Total	1,150	....	100.41	....

high stanines. This difference is expressed by the percent of student officers that exceed the mean of the group at each stanine level.

Dive Bombing and Skip Bombing

Only a few dive and skip bombing scores were collected, most of which were obtained in the Fourth Air Force. The product-moment correlations of these scores with pilot stanine are presented in table 2.37.

TABLE 2.37.—*Correlations of pilot stanine with dive and skip bombing scores*

P-38 PILOTS—FOURTH AIR FORCE

Criterion Battery		N	Stanine		Bombing score		r
			Mean	S. D.	Mean	S. D.	
Dive bombing:							
	1	[1]28	...	...	...	...	...
	2	286	6.37	1.65	49.32	9.88	*0.12
	Mixed	45	5.98	1.74	49.16	9.34	.13
	Total	359	6.28	1.66	49.23	9.85	*.10
Skip bombing:							
	1	[1]5	...	...	...	...	...
	2	74	6.02	1.70	49.02	11.13	.07
	Mixed	[1]12	...	...	...	...	...
	Total	91	5.96	1.73	49.48	10.65	.06

P-47 PILOTS—SECOND AIR FORCE

Criterion Battery	N	Stanine Mean	Stanine S. D.	Bombing score Mean	Bombing score S. D.	r
Dive bombing, mixed	64	6.70	1.69	3.84	1.61	−0.26
Skip bombing, mixed	64	6.70	1.69	66.0	19.5	.04

[1] Correlation coefficients were not determined from these samples.

Skeet Scores

Skeet scores were obtained for fighter pilots in training at two Second Air Force stations. The correlation between augmented pilot stanine and skeet score was 0.28** for 112 cases.

Mission Grades and Comments

One training station in the First Air Force and three stations in the Fourth Air Force systematically required their instructors to give each student an overall rating or grade on every mission flown. In addition to these grades instructors were encouraged to comment on the excellencies or faults observed in the flying of the missions. The correlations of the pilot stanine with the overall grades and with scores derived from the comments are shown in table 2.38. It is to be noted that the correlations of

TABLE 2.38.—*Relation of pilot stanine to mission card grades and comment scores*

AAF station	Grade		Comment scores	
	N	r	N	r
First SJ	196	0.05	387	**0.23
Fourth VNIII	90	.03	...	...
Fourth PV	26	.30	26	**.49
Fourth OV	105	.26**	105	**.37
Average r (by Fisher's z)	417	.12*	516	**.25

the stanines against the scores derived from comments are all significant at the one percent level.

Tower Reports or Mobile Control Violations

The number of violations of flying safety procedures or instances of gross errors in taxiing, take-offs, flying the traffic pattern and in landing offers a partially objective measure of basic flying skills. Unfortunately few stations in the Continental Air Forces maintained adequate records of these observations. The correlation between the numbers of errors recorded at two First Air Force Stations and the stanines of the pilots concerned was .14** for 552 pilots. This coefficient represents the average (by Fisher's *z*) of separate coefficients for different classes and stations.

A study was made in the Second Air Force of the landing discrepancies noted in the landings of student pilots during a sixty day period at Strother Field. Two types of scores were obtained for each pilot: percent of landings for which errors were noted; and average number of errors noted per landing made. In table 2.39 are given the correlations between these two types of scores and the stanines of the pilots involved.

TABLE 2.39.—*Correlations between landing discrepancies and stanines*

85 STUDENT PILOTS—SECOND AIR FORCE

Type of stanines	Stanines		Percent of landings with errors			Average number of errors per landing		
	Mean	S. D.	Mean	S. D.	r Sta.	Mean	S. D.	r Sta.
Augumented pilot	7.01	1.60	33.6	21	0.04	44.5	31	0.04
Navigator	5.87	1.89	33.6	21	.10	44.5	31	.01
Bombardier	6.31	1.98	33.6	21	.12	44.5	31	.04

Ratings

Four different types of ratings have been tried out with fighter pilots in the Continental Air Forces; ratings with a printed form, ratings secured through personal interviews with instructors, and rankings made by instructors and rankings by fellow students on each other. The first two types were secured in the First Air Force and the latter two in the Second Air Force. The correlation of stanines with ratings on the printed form are presented in table 2.40. In May 1945 ratings on the printed rating form were discontinued in the advanced phase of training in the First Air Force and the ratings secured by interviews with instructors substituted for them.

TABLE 2.40.—*Relation of pilot stanine to proficiency ratings in operational training*

P-47 PILOTS—FIRST AIR FORCE

Correlations between pilot stanine and:

Type of rating	N	Gunnery items	Flying items	Trait items	Total rating
Printed rating, basic stations	475	....	*0.12	*0.12	....
Printed rating, advanced stations	461	*0.11	.08	03	....
Interview rating, advanced stations	427	....	....	....	**0.19

The relation of pilot stanine to this type of proficiency rating is also shown in table 2.40 for a total of four commitments.

The results of validity studies of rankings by instructors and by fellow students in the Second Air Force are shown in table 2.41.

TABLE 2.41.—*Correlations between augmented pilot stanine and rankings by instructors and fellow students*

P-47 PILOTS—SECOND AIR FORCE

	N	Stanine, I		Ranking, I		r_b
		Mean	S. D.	Mean	S. D.	
Instructor rankings	73	6.63	1.63	50.6	25.6	0.06
Student rankings by:						
4 or more students	156	6.51	1.73	5.06	1.04	.07
8 or more students	31	6.20	1.86	5.07	1.00	.12

Flying Evaluation Board Reports

Among possible criteria of flying proficiency, appearance before a Flying Evaluation Board because of official reports of professional incompetence would logically be expected to indicate a low degree of proficiency. Whether because very few men entering the training air forces were inefficient fliers or whether officials were loath to remove a man from flying at this stage of the game, the fact was that relatively few student officers were ever brought before Flying Evaluation Boards. Three studies are reported here. Because of the small number of trainees re-evaluated, what little relationship was found between stanine and re-evaluation is not believed to be of much practical importance.

In table 2.42 are shown stanine comparisons for three groups of pilots appearing before Flying Evaluation Boards. This table includes all men on whom stanines could be found that met the First Air Force Flying Evaluation Boards between April 1944 and July 1945.

TABLE 2.42.—*Stanine comparisons for three groups of pilots appearing before flying evaluation boards*

NONPHYSICAL REASONS ONLY—FIRST AIR FORCE

Statistics	Flying deficiency	Personality	Unknown	All causes
Number of pilots	20	15	12	47
M (theoret.)	6.19	6.34	6.25	6.25
M (obtained)	5.60	6.27	5.92	5.89
S. D. (obtained)	1.82	1.49	0.90	1.52
S. E. (obtained M)	.42	.40	.27	.22

For comparison purposes it was assumed that the populations of pilots entering the First Air Force for operational training were random samples of pilots graduating from advanced schools in the AAF Training Command. A theoretical mean stanine of these populations was obtained by weighting the mean stanine for each Training Command class according to the proportions of that class found in the Flying Evaluation Board group. Comparisons were then made between the theoretical and obtained mean

stanines. In table 2.42 it can be seen that all groups of re-evaluated pilots had lower stanines than the theoretical mean of their population, this being particularly true of the men who met the Board because of flying deficiencies. Though the numbers involved are too small to warrant the use of critical ratios, the obtained differences can be roughly evaluated in terms of the data given in the table.

In the Second Air Force, of a total of 790 fighter pilots in the May and June 1944 commitments in the 72d Fighter Wing, 31 were re-evaluated by Flying Evaluation Board action. Another 127 had aircraft accidents, leaving 632 pilots whose records show they might be thought at least moderately successful in training. The biserial coefficient of correlation between pilot stanine and success or failure in training, as evidenced by elimination through FEB action was 0.09.

In the Fourth Air Force the stanines of 1894 P–38 fighter pilot trainees were studied in relation to re-evaluation. It was found that the lower the

TABLE 2.43.—*Relation between pilot stanine and reevaluation rate*

FOURTH AIR FORCE P–38 PILOTS

Pilot stanine	N_t	$N_{re\text{-}eval}$	Percent $_{re\text{-}eval}$
8, 9	322	8	2.5
6, 7	780	25	3.2
4, 5	660	23	3.5
1, 2, 3	132	7	5.3
Total	1,894	63	3.3

TABLE 2.44.—*Relation between pilot stanine and initial cause of FEB action*

FOURTH AIR FORCE P–38 PILOTS

Initial cause	N	M_{stan}	SD_{stan}	Probability
Pilot's own request	19	6.21	1.40	<0.01
Referral by instructor	27	4.67	1.79	

stanine, the greater the chances were that the trainee would be re-evaluated. The re-evaluation rate of the low stanine trainees was more than twice that of the high stanine trainees. These results are shown in table 2.43. The relationship between pilot stanine and initial cause of referral to the Flying Evaluation Board is shown in table 2.44.

Accidents

One of the most useful functions that the psychological program could perform would be to identify, and eliminate from training, accident prone individuals. Consequently, research personnel in all four of the Continental Air Forces devoted a considerable amount of time studying and analyzing accident records and correlating various accident factors with aptitude scores.

In the First Air Force the relationship between accidents and stanines are shown in tables 2.45 and 2.46. The accident group was matched with

a control non-accident population for: (1) Graduation from the same Training Command class. (2) Contemporary attendance at same basic and advanced stations. (3) Amount of flying experience. (4) Officer status in the Training Command (student officers or enlisted aviation cadets).

Examination of the data in table 2.45 indicates that there was a significant ($P<0.05$) difference between pilot stanines of pilots having accidents (all causes) and pilots in a control group (no accidents) and between pilots having accidents classified as caused by pilot error and pilots in the con-

TABLE 2.45.—*Relation of pilot stanine to aircraft accidents in various categories*

FIRST AIR FORCE

Category of accidents	N	Mean stanine	S. D.
Accident free (control)	204	6.55	1.52
Accidents (all causes)	344	6.28	1.61
Non-pilot-error accidents	115	6.37	1.57
Pilot-error (50 Percent or more)	204	6.25	1.63
Pilot-error (100 Percent)	285	6.32	1.60
Pilot-error (50–100 Percent)	19	5.58	1.74
Undetermined cause	25	6.12	1.64
One accident (pilot-error only)	172	6.27	1.60
Two accidents (both pilot-error)	10	6.50	1.18
Two accidents (at least one pilot-error)	13	6.15	1.25
Nonfatal (pilot-error)	185	6.26	1.59
Fatal (all causes)	38	6.13	1.98
Fatal (pilot-error)	19	6.16	2.04
Fatal (undetermined)	19	6.11	1.37

TABLE 2.46.—*Distribution by stanine of accident-free control and pilot-error accident pilots*

Stanine	Accident-free control		Pilot-error accidents	
	N	Per cent	N	Per cent
9	29	14.22	18	8.82
8	26	12.75	32	15.69
7	45	22.06	40	19.61
6	54	26.47	46	22.55
5	33	16.17	43	21.08
4	13	6.37	15	7.35
3	3	1.47	7	3.43
2	1	.49	2	.98
1	0	.00	1	.49
Total	204	100.00	204	100.00

trol group. None of the other differences between the stanines of the various groups was significant.

For the May and June commitments in the Second Air Force, the pilot stanines (with criteria) of 127 fighter pilots (P–47) who had had accidents were compared with the pilot stanines of 632 men who completed training without difficulty. The biserial correlation between stanines and accidents for the two groups was 0.09.**

Further data on Second Air Force fighter pilot accidents were available from a study of aircraft accidents of pilots in training from July 1944 to 1 January 1945. During this period, 342 fighter pilot trainees had aircraft accidents for 287 of whom stanines were found to be available. Comparisons between the mean pilot stanines (with criteria) of these pilots and

the mean stanines of two groups of controls without accidents showed a small difference in favor of the control groups. In table 2.47 are shown separate comparisons between the stanines of pilots with various categories of accidents and for pilots in the corresponding control groups. Control Group A represented pilots graduating from advanced schools at the same time and Control Group B represented a random group of single-engine pilots in the Second Air Force at the time of the accidents.

In a preliminary study in the Third Air Force the stanines of fighter pilots having aircraft accidents during the 2-month periods were compared with stanines of pilots in corresponding control groups. The contradictory results obtained in the two samples showed the importance of obtaining large samples and of controlling as many as possible of the factors involved. The results of a more extensive study in the Third Air Force are shown in table 2.48. Data were collected for all fighter aircraft accidents occurring

TABLE 2.47.—*Comparison of stanines of pilots with aircraft accidents and stanines of two control groups*

P-47 AND P-40 PILOTS—SECOND AIR FORCE

Origin of accidents	Number cases	Accident mean	Pilots S. D.	Control Group A		Control Group B	
				Mean	S. D.	Mean	S. D.
Pilot error	153	5.87	1.75	6.30	1.50	6.65	1.38
Unknown	40	6.30	1.70	6.27	1.53	6.67	1.43
Materiel and unavoidable	94	6.48	1.59	6.16	1.60	6.70	1.45

The differences between accident pilots and pilots in Control Group B are significant at the 20 percent level for accidents due to unknown and matériel and unavoidable reasons and at the 1 percent level for accidents due to pilot error.

between 1 June 1944 and 1 May 1945. The names of several thousand pilots trained in the Third Air Force during this same period were obtained from the Air Force Personnel Depots. From these names were chosen a control group in which date of training in the Air Force, Psychological Unit at which classification was carried out, and date of testing were controlled. Comparisons were made between the mean bombardier, navigator, pilot and pilot plus credit stanines for pilots having accidents ascribed to pilot error, accidents ascribed to materiel failure, accidents of causes undetermined, and stanines of the control group of pilots who had no aircraft accidents in operational training. As shown in the table, the differences were not large, but pilots with accidents ascribed to pilot error consistently had lower stanines than pilots of the control group or pilots who had accidents ascribed to materiel failure. These differences are significant at approximately the 10 percent level.

In the Fourth Air Force, accident data were collected on 1880 P-38 trainees, all of whom were in training during the last half of 1944 and the first quarter of 1945. A total of 280 trainees in this group had one or more accidents. Comparisons were made between the pilot stanines of the accident free group and those of groups of pilots having various types of

accidents. These data are presented in table 2.49. While the differences are not great, accident groups of all types except those for which the accident cause was unknown have lower mean pilot stanines than the accident free group.

Time Required to Complete Training

Individual stations in the Continental Air Forces have probably been more flexible in handling their training directives than have the stations and schools in the Training Command. Student officers are relatively on their own in completing the necessary fighter pilot requirements in operational training, and individuals within each commitment thus have some opportunity to finish the complete course of instruction in different lengths of time. A possible indication of individual differences in proficiency would

TABLE 2.48.—*Comparison of stanines of fighter pilots with various types of aircraft accidents and stanines of a control group*

THIRD AIR FORCE

Type of stanine	N=466, control pilots		Pilots with accidents ascribed to:					
			N=263, pilot error		N=134, materiel failure		N=81, undetermined	
	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.
Bombardier	6.3	1.7	6.0	1.7	6.2	1.7	6.3	1.6
Navigator	6.1	1.8	5.8	1.8	6.1	1.8	5.9	1.6
Pilot	6.3	1.6	6.0	1.6	6.2	1.6	6.6	1.2
Pilot plus credit	6.4	1.6	6.2	1.7	6.4	1.7	6.6	1.5

be the number of weeks or months required to complete such training. An exploratory study was done in the First Air Force of the relationship between stanine and time in months to complete training. The correlation between pilot stanine and time taken to complete basic training was 0.16.* That between pilot stanine and time taken to complete advanced training was 0.13. For total training (basic plus advanced) the correlation with pilot stanine was 0.21.** The coefficients for basic and for total training were based on data for 701 pilots while that for advanced training included 144 pilots.

Exploratory studies were also carried out in the Second Air Force to determine the relation between training accomplishment and the pilot stanines of fighter pilots in operational training. For the 25 pilots in the 2-28-45 class at Peterson Field it was found that the correlation between stanines and number of requirements completed at the end of eight weeks was 0.21, that the correlation between stanines and hours flown in operational training was —0.09, and that the correlation between stanines and number of requirements completed per hour flown was 0.42. On the basis of these encouraging results, a study was made of the December 1944 commitment of fighter pilots trained at three 72d Wing stations. Three scores representing training accomplished at the end of 5 weeks were obtained for each of the 75 pilots involved. These scores were the weighted number

Table 2.49.—*Distributions of pilot stanines by cause, number, and result of accidents*

FOURTH AIR FORCE

Pilot stanine	All cases	Cause of accident				Number of accidents				Result to personnel				Damage to aircraft			
						One		2 or More		Fatal		Not fatal		Little or some		Major or complete	
		Total Number	Not pilot Error	Cause Unknown	Pilot error	All causes	Pilot error	All causes	Pilot error	All causes	Pilot error	All causes	Pilot error	All causes	Pilot error	All causes	Pilot error
9	109	15	6	3	6	14	5	1	1	0	0	15	6	7	4	7	1
8	208	24	6	2	16	24	16	0	0	3	1	21	15	10	9	14	7
7	370	44	10	10	24	42	23	2	1	10	1	34	23	16	11	28	12
6	410	57	20	10	27	56	26	1	1	12	4	45	23	18	10	38	16
5	395	68	21	8	39	61	32	7	7	7	3	61	36	32	18	35	20
4	258	42	10	3	29	38	26	4	3	6	5	36	24	15	10	27	18
3	116	28	9	5	14	24	11	4	3	4	2	24	12	11	4	17	10
2	13	1	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1
1	1	1	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1
Total	1,880	280	82	41	157	161	141	19	16	42	16	233	141	109	66	168	86
Mean	5.90	5.61	5.66	5.85	5.53	5.67	5.60	4.84	4.87	5.64	5.00	5.61	5.59	5.45	5.86	5.37	5.23
S. D.	1.61	1.65	1.64	1.67	1.66	1.64	1.65	1.53	1.49	1.41	1.37	1.60	1.67	1.65	1.63	1.44	1.60

of missions reported (the first mission of a particular type was weighted more heavily than repetitions of that mission), the hours flown in operational training, and the number of AAF Training Standards completed. Correlations between these scores and the stanines of the pilots are shown in table 2.50. These coefficients represent the z average of the separate coefficients for each class and station. In contrast to the results obtained in the earlier study, the hours flown showed the greatest correlation with the pilot stanine.

TABLE 2.50.—*Correlations between stanines and training accomplished*[1]

75 P-47 PILOTS—SECOND AIR FORCE

All stations combined

Training accomplishment score	Type of stanine					
	Bombardier	Navigator	Pilot	Pilot+credit	Mean	S. D.
Weighted mission completed	0.18	0.08	0.16	*0.21	56.0	22
Hours flown	*.23	.19	**.31	*.26	42.5	18
AAF training standards completed	.07	[illegible]	.05	.18	13.8	7
Mean	6.32	[illegible]	6.48	6.67	.....	.....
S. D.	1.9	1.8	1.5	1.7	.....	.....

[1]Obtained by Fisher's z technique.

In interpreting the results from both Air Forces, mention should be made of the many variables such as weather, illness, maintenance and training irregularities which obviously influenced these scores. However, without denying that they are important, it is still difficult to see how such uncontrolled variables could have produced a spurious validity. Rather, it is surprising that statistically significant coefficients of validity are obtained at all when such factors were operating to an unknown extent.

Academic Grades

Research personnel of the Second Air Force were asked to assist in the evaluation of a fighter pilot proficiency test given in the 72d Fighter Wing. Data were available for a preliminary administration of the test to 46 student pilots at 2 stations. The correlations between the raw scores for each part and the total test and the stanines of the 40 pilots for whom stanines were available are given in table 2.51.

TABLE 2.51.—*Correlations between stanines and written proficiency test scores*

46 P-47 PILOTS—SECOND AIR FORCE

Section	Proficiency test		Correlation with stanines		
	Mean score	S. D.	B	N	P(aug.)
Unit A	150	26	0.2[illegible]	*0.29	−0.12
Unit B	26	6	.15	*.33	−.12
Unit C	127	22	*.28	**.44	.10
Unit D	11	5	.26	**.47	.09
Unit E	74	15	.21	**.49	.09
Unit F	41	8	.25	**.40	−.10
Unit G	170	16	*.33	*.29	.14
Unit H	38	20	.16	*.36	.23
Total test	687	84	*.30	**.45	.04

Quite an extensive study has been made on the relationship between stanines and various ground school grades of P-38 pilots in the Fourth Air Force. These results are presented in table 2.52.

TABLE 2.52—*Correlations between stanines and various ground school grades*

P-38 PILOTS—FOURTH AIR FORCE

Grade and station	Pilot stanine		Navigator stanine		Bombardier stanine		N	Criterion mean	S. D.
	N	r	N	r	N	r			
Armament:									
Santa Maria II	111	0.24	107	0.22	108	0.22	111	86.68	6.77
Santa Maria III	124	.06	124	−.03	124	.03	123	79.94	6.58
Engine Maintenance:									
Van Nuys III	96	.24	96	.36	96	.30	130	90.53	4.11
Santa Maria II	117	−.01	110	−.22	110	−.16	117	86.61	6.71
Santa Maria III	129	.16	128	.11	129	.09	128	87.30	6.72
Gunnery theory:									
Santa Maria II	112	.02	109	−.11	109	−.12	127	90.15	10.32
Santa Maria III	119	.11	119	.17	119	.24	119	90.06	8.42
Blinker I:									
Van Nuys III	96	.12	96	.19	96	.17	126	2.87	1.48
Blinker II:									
Van Nuys III	96	.25	93	.16	93	.25	126	4.10	2.00
Code:									
Van Nuys III	99	.37	96	.42	96	.39	117	4.01	2.98
Communications:									
Santa Maria II	111	.16	110	.24	108	.11	111	95.61	8.30
Meteorology:									
Santa Maria II	114	.22	111	.26	111	.18	114	88.08	4.18
Chemical Warfare:									
Santa Maria II	112	.13	109	.01	109	.07	[illegible]	76.49	4.38
Skeet:									
Santa Maria III	130	.20	130	.02	130	.19	130	55.19	11.88

Retention as Instructor

At the completion of their training a number of the fighter pilots in the First Air Force were retained as instructors for varying periods of time. Stanines were found for 209 of these. In order to determine the extent to which the pilot stanine was predictive of retention as instructor, the average stanine of these pilots was compared with the theoretical average stanine of a corresponding group of unselected pilots. The two averages were 6.46 and 6.05 respectively for instructors and unselected pilots. The theoretical stanine for the corresponding unselected group of pilots was obtained by computing a weighted average of the mean stanines of the advanced flying school classes of which these pilots were a part. This type of control is based upon the assumption that pilots entering operational training in the First Air Force are an unselected sample of pilots graduating from advanced schools.

Validation of Stanines Against Combinations of Criteria

Examination of the nature of various types of criterion scores, the low intercorrelations obtained between such scores and the opinions of supervisory personnel and officers returned from combat all lead to the conclusion that no single measure would adequately cover all aspects of combat flying. Although it is probably true that the ultimate aim of operational training was to enable the fighter pilot to hit a moving or stationary target, it was also believed important that the pilot be able to fly close formation, that he excel in "rat races," that he have no fear of flying "on the deck,"

that he be able to navigate with accuracy, etc. It is certain that the degree of proficiency shown in many of these activities would be highly correlated; but in the absence of accurate measures of such proficiencies it is impossible to determine at this time the degree to which they are related. In addition, the unreliability of such measures as are available makes it difficult to interpret the intercorrelations between them. Hence it seemed desirable to try out various combinations of criteria in validation studies although the optimum combination for this purpose was not known.

A number of studies of combined criteria were carried out in the First Air Force. In the preliminary study a standard-score method of combining criteria was used with the May and June 1944 commitments. Aerial gunnery scores, ground gunnery scores, printed proficiency ratings from basic training stations, and proficiency ratings secured from interviews with instructors at advanced training stations were combined into a total proficiency score. The correlations between the pilot stanines and the combined criterion scores are shown by station and commitment in table 2.53.

TABLE 2.53.—*Correlations of combined criterion scores with pilot stanine*

FIRST AIR FORCE

Commitment and station	N	Criterion score, x		Stanine, s		r_{xs}
		Mean	S. D.	Mean	S. D.	
May:						
Dover	43	149.72	15.12	6.84	1.31	**0.44
Millville	38	150.24	19.32	6.66	1.41	.17
Norfolk	39	149.55	16.35	6.41	1.32	.12
Suffolk	18	149.45	19.85	6.33	1.07	.25
June:						
Bluethenthal	34	150.90	18.90	6.26	1.54	**.46
Dover	29	145.60	16.70	6.17	1.42	*.45
Millville	29	149.75	18.60	6.45	1.67	−.13
Average (by Fisher's z)	230	......	......	......	......	**.28

Another way of expressing the relationship between pilot stanine and proficiency is to give the mean pilot stanine of groups of pilots of different levels of proficiency. From the distributions of the composite criterion scores for the pilots in the May and June 1944 commitments, the high and low group were selected by using as cut-off points one sigma distance above and below the means of the distributions. For the resulting high, middle, and low groups in each distribution the mean bombardier, navigator and pilot stanines were found. The combined results for the two commitments are shown in table 2.54.

A further study was done in the First Air Force in which an attempt was made to make the composite score include a still wider coverage of criteria of proficiency in operational training. In this connection, to save time and effort in connection with studies of *individual test validities,* it had been decided to carry out the initial validation of tests and stanines against smaller groups at the extremes of proficiency. The reasons for this decision are given in more detail in the section on Test Validation. However, since such groupings were being made for the test validation, stanine validities were also obtained on this basis.

The extreme high and low groups for the May, June, July, and August 1944 commitments were selected in the following manner: The high group was composed of all individuals who fell one-half sigma above the mean on the combined ground-aerial gunnery score and who were at least above the mean on the verbal interviews. From this group were eliminated all men who fell one and one-half sigma below the mean on the basic proficiency rating scale and the mission card comment scores, or who had been re-evaluated, or who had had an accident due to personnel failure. The low group was composed of all individuals who fell one and one-half sigma

TABLE 2.54.—*Mean stanines of high, middle and low proficiency groups*

FIRST AIR FORCE

May and June Commitments—(N = 230)

Proficiency group	N	Per cent	Mean bombardier stanine	Mean navigator stanine	Mean pilot stanine
High	41	17.7	6.8	6.1	7.1
Middle	148	64.3	6.3	5.9	6.5
Low	41	17.7	5.6	5.6	5.9
Total	230	100	6.3	5.9	6.5
High low difference	.....	.....	1.2	0.5	1.2

below the mean on the combined ground-aerial gunnery scores and who were also below the mean on the verbal proficiency rating. (In gunnery the cut off points for the high and low groups are not symmetrical since the distributions are markedly skewed in the direction of poor scores). From this group were excluded any individual who was one and one-half sigma above the mean on the basic rating scale and/or the mission card comment score. The number in the high, middle, and low groups, the mean pilot stanine and standard deviation for each group and the triserial correlation between various groups and pilot stanine are shown in table 2.55. The critical ratios of the differences between the stanines of each group are also shown in the table.

In the Second Air Force pilots suffering aircraft accidents and those eliminated from training by Flying Evaluation Board action were grouped together as a failure group in a study of stanine validity. The success group included all pilots without accidents or re-evaluation. The biserial co-

TABLE 2.55.—*Relationship of pilot stanine to high-low proficiency groupings*

MAY, JUNE, JULY, AUGUST COMMITMENTS—FIRST AIR FORCE

Means and Standard Deviations

	N	*M stanine*	*S.D.*
High	56	6.93	1.52
Middle	289	6.61	1.46
Low	66	5.98	1.35

Triserial $r =$ ** 0.20

Critical Ratios Among Means

High *v.* low	3.53
High *v.* middle	1.49
Middle *v.* low	3.24

efficients of correlation between pilot stanine (with credit) and success or failure in operational training is given in table 2.56 for pilots in the May and June 1944 commitments. Also shown are the separate validity coefficients (biserial) for accidents and Flying Evaluation Board action alone.

TABLE 2.56.—*Correlation of pilot stanine with success and failure in operational training*

SECOND AIR FORCE

Pilot stanine v.—	N[1]	N[2]	M[1]	M[2]	t	M	r
FEB elimination	632	31	6.00	5.67	1.61	5.98	0.09
Accidents	632	127	6.00	5.73	1.62	5.95	.09
FEB [illegible] nts	632	158	6.00	5.72	1 62	5.94	.10

[1] Pilots completing training without difficulty.
[2] Pilots having accidents or eliminated by FEB action.

In the Fourth Air Force the stanines of P–38 and P–39 fighter pilots of five categories of proficiency were compared. The five categories used represented various combinations of aircraft accidents and elimination by Flying Evaluation Board action. This study was based upon the records of all second lieutenants and flight officers in fighter pilot training in the Fourth Air Force from 1 January through November 1944 for whom stanines were available. The average stanines of pilots in the five groups are shown in table 2.57. It is apparent from the table that when the categories are arranged in a logical order from greatest to least proficiency, the average pilot stanine shows a similar progressive decrease with pilots in the least proficient category showing the lowest average pilot stanine.

The combination of two or more types of proficiency scores as a criterion against which to validate selection procedures has the logical advantage of providing better coverage of the total job of the combat pilot. Although it does not *of necessity* follow, one would rather expect such a combined criterion to result in higher validity coefficients than were obtained with particular proficiency measures when a general flying aptitude score such as the pilot stanine is studied. The data reported here on stanine validities

TABLE 2.57.—*Mean stanines of pilots of different categories of proficiency*

FOURTH AIR FORCE

	Mean pilot stanine	Standard deviation	Number of cases
Graduated without being involved in accidents	6.04	1.60	822
Graduated but involved in one pilot-error accident	5.96	1.73	76
Graduated but involved in more than one accident	5.41	1.65	17
Reevaluated and not involved in accidents	5.31	1.67	49
Reevaluated and involved in accidents	5.06	1.71	18

	Navigator stanine		Bombardier stanine		Number of cases
	Mean	S. D.	Mean	S. D.	
Graduated without being involved in accidents	5.87	2.00	5.93	1.84	822
Graduated but involved in one pilot-error accident	5.76	1.97	5.71	1.76	76
Graduated but involved in more than one accident	6.11	2.15	4.95	1.75	17
Reevaluated and not involved in accidents	4.75	2.04	5.00	1.71	49
Reevaluated and involved in accidents	5.83	1.53	5.22	1.75	18

against combinations of criteria do not support such an expectation. For example in table 2.53 the correlation between a combination of aerial and ground gunnery scores, ratings at basic schools and interview ratings at advanced schools yielded a validity coefficient of 0.28 for approximately 200 pilots. For the same pilots (May and June commitments only) the validity of the advanced ratings alone was 0.26. The 0.02 advantage for the combined criterion hardly warrants the extra computations involved. When still wider coverage of criteria was attempted as in table 2.55, the validity coefficient was 0.20 while that for the advanced school ratings alone on the same groups of pilots was .19 (table 2.40). Of course the interpretation of the data in table 2.55 is complicated by the fact that a triserial coefficient was used. However, there is certainly no clear evidence of any great superiority of the combined criteria. Similarly the combination of accidents and Flying Evaluation Board eliminations in table 2.57 raised the validity of the pilot stanine 0.01 from that obtained with each criterion separately. Insufficient data are available to determine whether or not this outcome is due to a high degree of intercorrelation between criteria or to other factors not yet recognized.

Miscellaneous Validation

Had they been available for study, a number of scores and ratings given pilots in the AAF Training Command could have been studied in relation to criteria of proficiency in operational training. For fighter pilots the only data of this type studied in the continental air forces were the fixed gunnery scores in the AAF Training Command. Data were available for 179 P–47 pilots in the First Air Force. The correlation between fixed gunnery scores in the AAF Training Command and scores in aerial and ground gunnery in operational training were 0.24 and 0.12 respectively. The correlation with the combined aerial and ground gunnery scores was 0.38. Although the number of cases is small, these coefficients indicate that gunnery performance in the Training Command is more closely related to gunnery performance in operational training than are stanines and classification test scores.

Test Validation

The relative effectiveness with which various criterion scores were predicted by the different classification tests is shown in tables 2.58 through 2.70 which follow. Data on the prediction of air-to-air gunnery scores are given for three groups of pilots: P–38, P–47 and P–61 pilots in tables 2.58—2.60. Correlations between test scores and air-to-ground gunnery scores for P–38 and P–61 pilots are given in table 2.61 and 2.62. The relation between test scores of P–38 pilots and skip bombing scores, dive bombing scores, mission grades, adversity scores and Flying Evaluation Board action are shown in tables 2.63—2.67. Finally, the degree of prediction of accidents by the test scores is shown in tables 2.68 and 2.69 for P–38 and P–61 pilots. The tables for P–38 pilots represents all available data on

pilots trained in the Fourth Air Force the last half of 1944 and the first half of 1945.

TABLE 2.58.—*Correlations between stanines and test scores (x) and air-to-air gunnery scores (y)*

P-38 PILOTS—FOURTH AIR FORCE

Predictor variables	M_x	SD_x	M_y	SD_y	N	r_{xy}
Bombardier stanine	5.96	1.79	49.73	10.01	1576	**.08
Navigator stanine	5.85	1.89	49.73	10.01	1576	*.06
Pilot stanine	5.93	1.62	49.73	10.01	1576	**.10
Rotary pursuit CM803A	52.95	9.57	49.43	10.13	903	*.07
Two hand coordination CM101A	53.66	9.59	49.79	10.04	1515	*.06
Complex coordination CM701A	53.47	9.55	49.73	10.03	1575	**.07
Aiming stress CM211A	47.19	10.74	49.79	10.04	1515	.08
Discrimination reaction time CP611D	54.37	8.71	49.73	10.01	1576	**.07
Finger dexterity CM116A	51.43	9.88	49.73	10.03	1575	.08
Reading comprehension C1614G	24.10	12.62	49.79	10.04	1515	.04
Mechanical principles C1903A	62.15	17.52	49.79	10.04	1515	.02
Mechanical information C1905A	15.86	8.52	49.49	10.23	802	.08
Dial and table reading CP622-21A	35.88	8.61	49.73	10.01	1576	**.07
Spatial orientation II CP503B	22.92	8.62	49.73	10.01	1576	*.05
Spatial orientation I CP501B	28.80	5.69	49.73	10.01	1576	.04
Numerical operations (F) C1702B	34.99	11.86	49.73	10.01	1576	.02
Numerical operations (B) C1702B	32.25	11.54	49.73	10.01	1576	—.01
Mathematics A C1702E	26.64	14.49	49.39	10.25	863	.04
Mathematics A C1702F	25.14	16.09	50.13	9.75	713	.05
Mathematics B C1710A	39.45	15.86	49.39	10.25	363	.06
Mathematics B C1206C	15.77	9.07	50.13	9.75	713	—.03
Biographical data (P) CE602D	29.22	6.28	50.17	9.90	636	.05
Biographical data (N) CE602D	22.48	3.17	50.17	9.90	636	.04
Speed of identification CP610A	34.32	7.18	49.73	10.01	1576	**.09
General information (N) CE505D	21.21	5.52	50.13	9.75	713	.07
General information (P) CE505D	36.46	5.94	50.13	9.75	713	.01
Technical vocabulary (P) CE505C	21.22	6.69	49.39	10.23	863	—.02
Technical vocabulary (B) CE505C	5.60	3.19	49.39	10.25	863	.01
Technical vocabulary (N) CE505C	15.10	8.20	49.39	10.25	863	.04
Officer quality	44.70	9.19	50.17	9.90	636	.05

TABLE 2.59.—*Correlations between test scores (x) and air-to-air gunnery scores (y)*

P-47 PILOTS—SECOND AIR FORCE

Test variables	M_x	SD_x	M_y	SD_y	N	r_{xy}
Rotary pursuit CM 803A	51.24	8.16	[illegible].80	7.03	90	*.22
Rotary pursuit (Div.Att.) CM 410B	52.87	9.85	9.08	6.45	257	.05
Two hand coordination CM 101A	53.73	9.40	8.90	6.79	357	*.10
Complex coordination CM701A	54.87	10.23	8.94	6.84	358	.08
Aiming stress CE 211A	52.52	10.51	8.39	6.66	336	.00
Discrimination reaction time CP611D	55.62	7.31	8.94	6.84	358	*.12
Finger dexterity CM116A	51.27	9.87	8.94	6.84	358	.02
Reading comprehension CI614G	22.87	10.57	8.63	6.78	337	—.02
Reading comprehension AC10D	42.45	13.08	15.00	9.79	11	.19
Reading comprehension C1614H	27.90	12.40	13.65	2.64	11	.17
Mechanical principles C1903A	63.19	16.64	8.63	6.78	337	*.12
Mechanical principles C1903B	41.46	11.49	13.65	2.64	11	*.44
Mechanical information C1905A	14.33	7.82	7.97	7.45	91	.11
Mechanical comprehension AC10D	12.68	6.23	15.00	9.79	11	*.[illegible]
Dial and Table Reading CP622-21A	37.19	7.74	8.79	6.74	348	.05
Dial reading CP622A	26.09	5.76	15.00	9.79	11	.19
Spatial orientation II CP503B	22.22	6.58	8.96	6.97	359	—.07
Spatial orientation I CP501B	29.07	5.44	8.96	6.97	359	—.04
Numerical operations I CI702B	35.38	10.54	8.88	6.49	246	—.10
Numerical operations II CI702B	30.77	11.33	8.88	6.49	246	—.16
Mathematics A CI 702E	27.76	17.30	8.73	8.04	102	.08
Mathematics A CI 702F	23.86	14.78	8.91	6.37	255	.04
Mathematics B CI 710A	38.68	15.93	7.97	7.45	91	.18
Mathematics B CI 206C	16.46	9.09	9.08	6.45	257	—.01
Biographical data (P) CE 602D	30.00	6.17	9.08	6.45	257	—.06
Biographical data (N) CE 602D	22.57	3.13	9.08	6.45	257	—.03
Speed of identification CP 610A	46.91	4.39	8.63	6.78	337	.03
General information (N) CE 505D	20.94	5.29	8.38	5.49	246	.06
General information (P) CE 505D	37.17	6.21	8.38	6.49	246	.02
General information CE 505E	52.90	11.65	13.65	2.64	11	.29
Technical vocabulary (P) CE 505C	20.50	7.56	8.73	8.04	102	.01
Technical vocabulary (B) CE 505C	5.67	3.23	8.73	8.04	102	.06
Technical vocabulary (N) CE 505C	15.27	8.46	8.73	8.04	102	—.02
Officer quality	45.75	8.76	8.73	6.51	251	.03

TABLE 2.60.—*Correlations between stanines and test scores (x) and air-to-air Gunnery Scores (y)*

P-61 PILOTS—FOURTH AIR FORCE

Predictor variables	M_x	SD_x	M_y	SD_y	N	r_{xy}
Bombardier stanine	49.66	9.87	6.26	1.75	204	.12
Navigator stanine	49.64	9.87	6.21	1.80	204	.08
Pilot stanine	49.66	9.87	6.18	1.61	204	**.10
Rotary pursuit CM 803A	49.52	9.64	52.32	9.55	195	.08
Two hand coordination CM 101A	49.60	9.72	53.36	37.08	197	.08
Complex coordination CM701A	49.66	9.87	55.16	16.77	204	*.16
Finger dexterity CM116A	49.66	9.87	52.02	9.15	204	.10
Discrimination reaction time CP611D	49.56	9.83	54.92	9.70	203	.06
Aiming stress CE 211A	49.73	9.53	48.05	10.12	192	.07
Technical vocabulary (P) CE505C	49.14	9.44	20.96	6.46	169	.09
Technical vocabulary (N) CE505C	49.14	9.44	15.79	8.38	169	.07
Technical vocabulary (B) CE505C	49.14	9.44	5.77	3.24	169	.08
Speed of identification CP610A	49.73	9.53	17.63	3.43	192	.08
Mathematics A C1702E	49.14	9.44	28.85	17.12	169	.08
Mathematics A C1702F	51.05	10.92	25.64	14.45	55	.09
Mathematics B C1710A	49.20	8.95	41.18	14.97	137	*.16
Mathematics B C1206C	51.05	10.92	19.0C	8.74	55	.17
Reading comprehension C1614G	49.73	9.53	25.99	12.10	192	.08
Mechanical principles C1903A	49.73	9.53	62.70	15.86	192	*.14
Mechanical information C1905A	49.20	8.85	15.12	7.68	137	.07
Dial and table reading CP621-22A	49.73	9.53	36.48	8.81	192	*.16
Numerical operations I C1702B	49.66	9.87	17.18	5.89	204	.07
Numerical operations II C1702B	49.66	9.87	15.94	5.56	204	.07
Spatial orientation I CP501B	49.66	9.87	59.12	12.55	204	−.09
Spatial orientation II CP503B	49.66	9.87	23.86	6.52	204	.09
General information (P) CE505D	51.05	10.92	37.55	5.46	55	−.14
General information (N) CE505D	51.05	10.92	22.96	4.78	55	*−.25
Officer quality score	51.64	9.96	47.45	9.83	38	−.12

TABLE 2.61.—*Correlations between stanines and test scores (x) and air-to-ground gunnery scores (y)*

F-38 PILOTS—FOURTH AIR FORCE

Predictor variables	M_x	SD_x	M_y	SD_y	N	r_{xy}
Bombardier stanine	5.89	1.78	49.92	9.64	1274	**.08
Navigator stanine	5.77	1.86	49.92	9.64	1274	*.07
Pilot stanine	5.87	1.60	49.92	9.64	1274	**.09
Rotary pursuit CM903A	52.76	9.48	49.75	9.28	659	.01
Two hand coordination CM101A	53.52	9.49	50.04	9.54	1236	**.08
Complex coordination CM701A	53.20	9.33	49.92	9.66	1273	.03
Aiming stress CM211A	46.[illegible]	10.68	50.04	9.54	1236	.04
Discrimination reaction time CP611D	54.18	8.73	49.92	9.64	1274	.04
Finger dexterity CM116A	51.11	9.75	49.92	9.64	1274	.02
Reading comprehension C1614G	23.34	12.29	50.04	9.54	1236	*.06
Mechanical principles C1903A	62.02	17.72	50.04	9.54	1236	**.08
Mechanical information C1905A	15.70	7.91	49.52	9.04	586	.06
Dial and table reading CP622-21A	35.78	8.27	49.92	9.64	1274	.03
Spatial orientation II CP503B	22.80	6.70	49.92	9.64	1274	.05
Spatial orientation I CP501B	28.61	5.74	49.92	9.64	1274	.01
Numerical operations (F) C1702B	35.01	11.78	49.92	9.64	1274	.02
Numerical operations (B) C1702B	32.21	11.37	49.92	9.64	1274	.02
Mathematics A C1702E	25.75	17.20	49.32	9.19	624	*.08
Mathematics A C1702F	24.78	16.12	50.50	10.00	650	.05
Mathematics B C1710A	38.80	15.67	49.32	9.19	624	.05
Mathematics B C1206C	15.74	8.83	50.50	10.00	650	.05
Biographical data (P) CE602D	29.15	6.28	50.36	10.02	595	.01
Biographical data (N) CE602D	22.47	3.17	50.36	10.02	595	−.03
Speed of identification CP610A	34.21	7.28	49.92	9.64	1274	.01
General information (N) CE505D	21.05	5.48	50.50	10.00	650	.04
General information (P) CE505D	36.35	6.03	50.50	10.00	650	.02
Technical vocabulary (P) CE505C	21.07	6.59	49.32	9.19	624	.06
Technical vocabulary (B) CE505C	5.51	3.11	49.32	9.19	624	−.01
Technical vocabulary (N) CE505C	14.67	8.09	49.32	9.19	624	.05
Officer quality	44.62	9.16	50.36	10.02	595	.07

TABLE 2.62.—*Correlations between stanines and test scores (x) and air-to-ground gunnery scores (y)*

P-61 PILOTS—FOURTH AIR FORCE

Test variables	M_x	SD_x	M_y	SD_y	N	r_{xy}
Bombardier stanine	49.69	9.86	6.14	1.74	246	.08
Navigator stanine	49.69	9.86	6.09	1.85	246	.09
Pilot stanine	49.69	9.86	6.17	1.62	246	.07
Rotary pursuit CM803A	49.59	9.77	51.92	9.34	231	.04
Two Hand Coordination CM101A	49.63	9.81	52.74	9.59	235	.08
Complex coordination CM701A	49.69	9.86	54.81	10.19	246	.06
Finger dexterity CM116A	49.69	9.86	52.05	9.33	246	*.13
Discrimination reaction time CP611D	49.69	9.86	54.72	9.51	246	−.06
Steadiness under pressure CE206B	48.86	9.51	54.29	5.25	42	.07
Aiming stress CE211A	49.76	9.83	47.65	10.02	202	.06
Technical vocabulary (P) CE 505C	49.37	9.67	20.72	6.74	190	.10
Technical vocabulary (N) CE 505C	49.37	9.67	15.18	8.24	190	*.18
Technical vocabulary (B) CE 505C	49.37	9.67	5.76	3.26	190	.11
Speed of identification CP610A	49.76	9.83	17.58	3.40	202	−.06
Mathematics A C1702E	49.37	9.67	27.65	17.12	198	.06
Mathematics A C1702F	50.79	10.31	25.30	14.50	56	.02
Mathematics B C1710A	49.36	9.64	40.86	14.99	146	.00
Mathematics B C1206C	50.79	10.31	19.04	8.65	56	.03
Reading comprehension AC10D-II	49.39	9.85	36.82	10.25	44	.20
Reading comprehension C1614G	49.76	9.83	25.49	11.91	202	.06
Mechanical comprehension AC10D-VI	49.39	9.85	9.99	5.82	44	*.37
Mechanical principles C1903A	49.76	9.83	62.71	10.29	202	.13
Mechanical information C1905A	49.36	9.64	15.20	7.69	146	.08
Dial and table reading CP621-22A	49.76	9.83	36.02	8.81	202	−.06
Numerical operations I C1702B	49.69	9.86	17.02	5.77	246	*−.13
Numerical Operations II C1702B	49.69	9.86	15.93	5.74	246	*−.16
Spatial orientation I CP501B	49.69	9.86	59.20	12.22	246	−.01
Spatial orientation II CP503B	49.69	9.86	23.70	6.40	246	−.01
Biographical data (P) CE602D	51.67	9.49	31.79	6.37	39	−.01
Biographical data (N) CE602D	51.67	9.49	23.13	3.87	39	.06
General information (P) CE505D	50.79	10.31	37.66	5.54	56	−.20
General Information (N) CE505D	50.79	10.31	22.75	4.68	56	−.19
Numerical approximation C1706A	49.39	9.85	11.05	6.04	44	.11
Arithematic reasoning CI 706A	49.39	9.85	13.34	5.33	44	.21
Table reading CP 621A	49.39	9.85	45.59	10.60	44	−.19
Dial reading CP622A	49.39	9.85	26.70	7.33	44	−.16
Officer quality score	51.67	9.49	47.51	9.76	39	.09

TABLE 2.63.—*Correlations between stanines and test scores (x) and skip bombing scores (y)*

P-38 PILOTS—FOURTH AIR FORCE

Predictor variables	M_x	SD_x	M_y	SD_y	N	r_{xy}
Bombardier stanine	5.99	1.67	50.19	10.19	462	.06
Navigator stanine	5.79	1.71	50.19	10.19	462	.03
Pilot stanine	5.82	1.57	50.19	10.19	462	*.11
Rotary pursuit CM803A	51.28	11.06	50.08	10.21	444	*.12
Two hand coordination CM101A	54.26	9.20	50.24	9.94	456	−.01
Complex coordination CM701A	52.58	8.79	50.19	10.19	462	.10
Aiming stress CM211A	46.32	10.39	50.24	9.94	456	−.01
Discrimination reaction time CP611D	55.05	8.07	50.19	10.19	462	.06
Finger dexterity CM116A	51.06	9.99	50.19	10.19	462	.06
Reading comprehension C1614G	24.13	11.75	50.24	9.94	456	.05
Mechanical principles C1903A	64.64	17.99	50.24	9.94	456	*.11
Mechanical information C1905A	14.87	8.09	49.61	11.00	70	.08
Dial and table reading CP622-21A	36.08	7.66	50.19	10.19	462	.01
Spatial orientation II CP503B	22.38	6.46	50.19	10.19	462	.05
Spatial orientation I CP501B	28.32	5.77	50.19	10.19	462	.06
Numerical operations (F) C1702B	34.81	11.78	50.19	10.19	462	−.06
Numerical operations (B) C1702B	32.56	11.01	50.19	10.19	462	−.02
Mathematics A C1702E	25.05	15.55	49.34	12.26	76	.02
Mathematics A C1702F	25.49	16.95	50.36	9.70	386	.04
Mathematics B C1710A	38.57	15.91	49.34	12.26	76	.06
Mathematics B C1206C	16.02	9.32	50.36	9.70	386	−.01
Biographical data (P) CE602D	29.06	6.25	50.37	9.81	374	.07
Biographical data (N) CE602D	22.29	3.20	50.37	9.81	374	.01
Speed of identification CP610A	34.52	6.97	50.19	10.19	462	.05
General information (N) CE505D	21.23	5.39	50.36	9.70	386	.09
General information (P) CE505D	36.56	6.35	50.36	9.70	386	.05
Technical vocabulary (P) CE505C	21.47	6.94	49.34	12.26	76	−.05
Technical vocabulary (B) CE505C	5.75	2.85	49.34	12.26	76	−.07
Technical vocabulary (N) CE505C	15.96	8.09	49.34	12.26	76	.11
Officer quality	45.57	9.99	50.37	9.81	374	.06

Table 2.64.—*Correlations between stanines and test scores (x) and dive bombing scores (y)*

P-38 PILOTS—FOURTH AIR FORCE

Predictor variables	M_x	SD_x	M_y	SD_y	N	r_{xy}
Bombardier stanine	6.07	1.78	49.36	9.97	822	.03
Navigator stanine	5.92	1.82	49.36	9.97	822	−.01
Pilot stanine	5.95	1.61	49.36	9.97	822	.05
Rotary pursuit CM803A	52.61	9.78	49.20	10.06	342	.02
Two hand coordination CM101A	54.38	9.43	49.40	9.96	795	*.07
Complex coordination CM701A	53.52	9.32	49.36	9.97	822	.06
Aiming stress CM211A	46.82	10.61	49.40	9.96	795	.03
Discrimination reaction time CP611D	55.00	8.62	49.36	9.97	822	.03
Finger dexterity CM116A	51.21	9.99	49.36	9.97	822	.04
Reading comprehension C1616G	34.85	12.23	49.40	9.96	795	.01
Mechanical principles C1903A	63.34	17.88	49.40	9.96	795	.03
Mechanical information C1905A	15.95	8.00	49.28	10.26	285	.07
Dial and table reading CP622-21A	36.29	8.01	49.36	9.97	822	−.03
Spatial orientation II CP503B	23.04	6.51	49.36	9.97	822	−.03
Spatial orientation I CP501B	28.77	5.78	49.36	9.97	822	−.06
Numerical operations (F) C1702B	35.08	11.78	49.36	9.97	822	.03
Numerical operations (B) C1702B	32.76	11.30	49.36	9.97	822	.00
Mathematics A C1702E	28.93	18.02	49.19	10.25	312	.0[illegible]
Mathematics A C1702F	25.19	16.70	49.47	9.77	510	.03
Mathematics B C1710A	40.12	15.64	49.19	10.25	312	−.03
Mathematics B C1206C	16.12	9.16	49.47	9.77	510	−.01
Biographical data (P) CE602D	29.08	6.28	49.56	9.80	465	.05
Biographical data (N) CE602D	22.35	3.14	49.56	9.80	465	.01
Speed of identification CP610A	34.62	7.07	49.36	9.97	822	.02
General information (N) CE505D	21.32	5.42	49.47	9.77	510	.03
General information (P) CE505D	36.77	6.16	49.47	9.77	510	.04
Technical vocabulary (P) CE505C	21.46	6.47	49.19	10.25	312	.05
Technical vocabulary (B) CE505C	5.94	3.19	49.19	10.25	312	.00
Technical vocabulary (N) CE505C	16.02	8.45	49.19	10.25	312	.03
Officer quality	45.76	9.40	49.56	9.80	465	−.01

Table 2.65.—*Correlations between stanines and test scores (x) and mission grades (y)*

P-38 PILOTS—FOURTH AIR FORCE

Predictor variables	M_x	SD_x	M_y	SD_y	N	r_{xy}
Bombardier stanine	5.97	1.77	49.33	10.57	432	.06
Navigator stanine	5.80	1.77	49.33	10.57	432	.05
Pilot stanine	5.75	1.56	49.33	10.57	432	*.11
Rotary pursuit CM803A	51.55	10.94	49.24	10.64	415	**.16
Two hand coordination CM101A	53.42	9.60	49.40	10.35	420	.09
Complex coordination CM701A	52.24	9.30	49.33	10.57	432	.01
Aiming stress CM211A	46.47	10.77	49.40	10.35	420	.06
Discrimination reaction time CP611D	55.66	7.30	49.33	10.57	432	−.03
Finger dexterity CM116A	51.34	9.92	49.33	10.57	432	−.07
Reading comprehension C1614G	24.35	11.89	49.40	10.35	420	.09
Mechanical principles C1903A	64.27	16.95	49.40	10.35	420	*.10
Mechanical information C1905A	15.44	8.00	48.[illegible]	9.30	84	−.05
Dial and table reading CP622-21A	35.63	8.39	49.33	10.57	432	−.02
Spatial orientation II CP503B	22.25	6.62	49.33	10.57	432	.04
Spatial orientation I CP501B	27.97	5.82	49.33	10.57	432	.01
Numerical operations (F) C1702B	34.67	12.49	49.33	10.[illegible]	432	.07
Numerical operations (B) C1702B	32.64	11.48	49.33	10.57	432	.00
Mathematics A C1702E	28.78	17.31	48.42	10.47	96	.14
Mathematics A C1702F	25.62	16.47	49.59	10.59	[illegible]	.06
Mathematics B C1710A	40.31	17.41	48.42	10.47	96	−.03
Mathematics B C1206C	15.57	8.65	49.59	10.59	336	.09
Biographical data (P) CE602D	29.00	6.56	49.48	10.53	314	.10
Biographical data (N) CE602D	22.28	3.19	49.48	10.53	314	.01
Speed of identification CP610A	34.06	7.03	49.33	10.57	432	−.[illegible]
General information (N) CE505D	21.36	5.18	49.59	10.59	336	.0[illegible]
General information (P) CE505D	35.98	6.19	49.59	10.59	336	*.13
Technical vocabulary (P) CE505C	21.90	5.80	48.42	10.47	96	*−.20
Technical vocabulary (B) CE505C	5.83	3.55	48.42	10.47	96	.03
Technical vocabulary (N) CE505C	15.80	8.63	48.42	10.47	96	.11
Officer quality	45.71	9.09	49.48	10.53	314	**.18

TABLE 2.66.—*Correlations between stanines and test scores (x) and adversity scores (y)*

P-38 PILOTS—FOURTH AIR FORCE

Predictor variables	M_x	SD_x	M_y	SD_y	N	r_{xy}
Bombardier stanine	5.92	1.70	50.73	10.75	133	**.25
Navigator stanine	5.76	1.63	50.73	10.75	133	*.20
Pilot stanine	5.71	1.53	50.73	10.75	133	**.28
Rotary pursuit CM803A	50.66	11.53	50.51	10.77	130	**.26
Two hand coordination CM101A	52.16	9.16	50.73	10.75	133	*.17
Complex coordination CM701A	51.60	8.82	50.73	10.75	133	.15
Aiming stress CM211A	45.95	11.47	50.73	10.75	133	.06
Discrimination reaction time CP611D	56.17	6.99	50.73	10.75	133	.10
Finger dexterity CM116A	50.97	9.55	50.73	10.75	133	.08
Reading comprehension C1614G	22.37	11.21	50.73	10.75	133	*.19
Mechanical principles C1903A	64.17	16.00	50.73	10.75	133	.14
Dial and table reading CP622-21A	35.42	8.20	50.73	10.75	133	.14
Spatial orientation II CP503B	22.00	6.88	50.73	10.75	133	.08
Spatial orientation I CP501B	27.66	6.00	50.73	10.75	133	.06
Numerical operations (F) C1702B	35.63	12.69	50.73	10.75	133	.11
Numerical operations (B) C1702B	32.96	10.73	50.73	10.75	133	.12
Mathematics A C1702F	23.91	14.83	50.55	10.81	115	.13
Mathematics B C1206C	14.44	8.19	50.55	10.81	115	.17
Biographical Data (P) CE602D	28.95	5.97	50.55	10.81	115	−.01
Biographical data (N) CE602D	21.86	3.16	50.55	10.81	115	−.08
Speed of identification CP610A	33.80	6.88	50.73	10.75	133	.08
General information (N) CE505D	20.48	5.01	50.55	10.81	115	.06
General information (P) CE505D	35.16	6.59	50.55	10.81	115	*.22
Officer quality	43.80	7.44	50.55	10.81	115	**.28

TABLE 2.67.—*Stanines and test scores of pilots that completed training and pilots re-evaluated for flying deficiency by FEB boards*

P-38 PILOTS—FOURTH AIR FORCE

Test variable	Re-evaluated			Not re-evaluated			Diff.	SE	P
	Mean	S. D.	N	Mean	S. D.	N			
B stanine	5.27	1.72	59	5.96	1.79	1576	0.69	0.23	<0.01
N stanine	5.19	1.82	59	5.85	1.89	1576	.66	.24	<.01
P stanine	5.41	1.62	59	5.93	1.62	1576	.52	.21	<.05
Rotary pursuit CM803A	49.71	9.55	53	52.40	10.01	1418	2.69	1.34	<.05
Two hand coordination	51.60	8.44	56	53.64	9.57	1438	2.04	1.15	<.05
Complex coordination CM701A	52.91	8.88	58	53.47	[illegible]	1575	.56	1.19	......
Aiming stress CM211A	48.48	11.60	58	47.19	10.74	1515	−1.27	1.54	......
Discrimination reaction time CP611D	51.78	11.20	58	54.37	8.71	1576	2.59	1.48	<.05
Finger dexterity CM116A	51.69	9.03	59	51.43	9.88	1575	− .26	1.20	......
Technical vocabulary (P) CE505C	22.19	6.84	48	21.22	6.69	863	− .97	1.01	......
Technical vocabulary (N) CE505C	12.88	7.35	48	15.10	8.20	863	2.22	1.10	<.05
Technical vocabulary (B) CE505C	5.92	3.00	48	5.60	3.19	863	− .32	.14	<.05
Speed of identification CP610A	32.32	8.41	59	34.32	7.13	1576	2.00	1.11	<.05
Mathematics A C1702E	18.69	16.04	48	26.64	14.49	863	7.95	2.37	......
Mathematics B C1710A	36.56	12.70	48	39.45	15.86	862	2.89	1.91	......
Mechanical information C1905A	14.15	7.86	39	15.86	8.52	802	1.71	1.29	......
Dial and table reading CP622-21A	34.53	9.61	39	35.88	8.61	1576	1.35	1.55	......
Numerical operations (F) C1702B	34.31	12.90	59	34.99	11.86	1576	.68	1.71	......
Numerical operations (B) C1702B	29.36	12.07	59	32.25	11.54	1576	2.89	1.59	<.05
Spatial orientation I CP501B	28.02	5.53	59	28.80	5.69	1576	.78	.23	<.01
Spatial orientation II CP503B	21.59	7.41	59	22.29	8.62	1576	.70	.31	<.05
Mechanical principles CI903A	55.88	18.20	50	62.15	17.52	1515	6.27	2.60	<.05
Reading comprehension C1614G	19.24	13.22	50	24.10	12.62	1515	4.86	1.48	<.01

TABLE 2.68.—*Correlations between stanines and test scores and occurrence of accidents ascribed to pilot error*

P-38 PILOTS—FOURTH AIR FORCE

o = Pilots completing training without accidents. t = Pilots with accidents ascribed to pilot error.

Predictor Variables	N_t	P_o	M_o	M_t	SD_t	r_{bis}
Bombardier stanine	1721	0.902	5.95	5.80	1.77	0.0[illegible]
Navigator stanine	1721	.902	5.83	5.80	1.87	.0[illegible]
Pilot stanine	1721	.902	5.93	5.59	1.60	**.11
Two hand coordination CM101A	1653	.906	53.75	52.20	9.63	*.0[illegible]
Complex coordination CM701A	1716	.903	53.34	51.91	9.44	*.0[illegible]
Aiming stress CM211A	1653	.906	47.16	46.75	10.68	.0[illegible]
Discrimination reaction time CP611D	1716	.903	54.43	53.89	8.60	.0[illegible]
Finger dexterity CM116A	1717	.903	51.38	50.56	9.96	.0[illegible]
Reading comprehension C1614G	1654	.905	24.17	23.38	12.50	.03
Mechanical principles C1903A	1656	.905	62.65	59.97	17.49	*.08
Dial and table reading CP622-21A	1718	.903	35.76	35.74	8.56	.06
Spatial orientation II CP503B	1718	.902	22.88	21.97	6.65	.07
Spatial orientation I CP501B	1717	.903	28.58	28.98	5.72	−.03
Numerical operations (F) C1702B	1718	.903	34.75	36.35	11.69	−.07
Numerical operations (B) C1702B	1718	.903	32.29	32.12	11.50	[illegible]
Mathematics A C1702F	837	.921	25.04	24.56	16.27	[illegible]
Mathematics B C1206C	837	.921	15.80	15.45	9.20	.02
Mathematics A C1702E	883	.884	26.50	26.84	17.53	−.01
Mathematics B C1710A	883	.885	39.31	38.89	15.82	.01
Biographical data (P) CE602D	755	.917	29.47	28.49	6.41	*.08
Biographical data (N) CE602D	755	.917	22.43	22.33	3.13	.02
Speed of identification CP610A	1719	.902	34.27	34.11	7.27	.01
General information (N) CE505D	837	.921	21.19	21.67	5.39	−.06
General information (P) CE505D	837	.921	36.58	36.36	5.14	.0[illegible]
Technical vocabulary (P) CE505C	884	.885	21.23	21.01	6.58	.02
Technical vocabulary (B) CE505C	884	.885	5.60	5.29	3.21	[illegible]
Technical vocabulary (N) CE505C	884	.885	15.10	14.35	8.16	.06
Mechanical information C1905A	819	.889	15.76	14.92	8.05	.06
Officer quality	753	.918	44.97	44.63	9.56	.02

TABLE 2.69.—*Correlations between stanines and test scores and occurrence of accidents ascribed to pilot error*

P-61 PILOTS—FOURTH AIR FORCE

e = Pilots with accidents ascribed to pilot error. t = All pilots (including those with accidents).

Predictor variables	N_t	N_e	M_t	M_e	SD_t	r_{bis}
Bombardier stanine	336	39	6.14	6.08	1.68	.03
Navigator stanine	336	39	5.96	6.23	1.68	−.08
Pilot stanine	336	39	6.16	5.87	1.66	.10
Rotary pursuit CM 803A	277	37	51.90	49.78	9.28	.14
Two hand coordination CM101A	283	37	53.18	55.19	9.68	.00
Complex coordination CM701A	333	36	56.19	58.78	8.31	**−.25
Aiming stress CE211A	244	35	47.53	46.40	10.41	.07
Discrimination reaction time CP611D	335	39	54.50	54.54	9.26	.00
Finger dexterity CM116A	336	39	52.45	52.23	9.69	.01
Technical vocabulary (P) CE505C	266	25	20.79	19.76	6.82	.08
Technical vocabulary (B) CE505C	266	25	5.94	5.44	3.28	.08
Technical vocabulary (N) CE505C	266	25	14.94	17.28	8.26	−.16
Speed of identification CP610A	292	39	17.49	16.79	3.45	.12
Mechanical principles CI903A	244	35	62.14	59.94	16.12	.09
Mathematics A C1702E	266	25	26.12	33.20	16.80	*−.23
Mathematics B C1710A	222	25	40.36	45.72	15.01	−.21
Spatial orientation I CP501B	336	39	59.23	57.59	12.06	.06
Spatial orientation II CP503B	336	39	23.73	23.64	6.94	.01
Reading comprehension CI614G	244	35	25.50	25.54	11.87	.00
Mechanical information CI905A	174	21	15.09	11.76	7.91	*.25
Numerical operations I C1702B	336	39	15.30	17.79	5.81	−.05
Numerical operations II C1702B	336	39	16.08	17.36	5.66	−.14
Dial and table reading CP622-21A	292	39	35.78	36.28	[illegible]	−.04

From the data in the above tables a number of observations can be made. In general, it is clear that no one test showed a high degree of relationship with any of the criteria studied. Although of statistical significance, no correlation between a single test and a criterion is sufficiently high to be of practical value in prediction. For convenience in obtaining an over-all view of test validity there are listed in table 2.70 all validity coefficients for P–38 pilots that were sufficiently large to be of statistical significance. From this table it is clear that the test battery as a whole best predicted the criterion of elimination by Flying Evaluation Board action with prediction of air-to-air gunnery performance a close second. Among the various predictor variables the stanines, being based upon combinations of test scores, naturally predicted the largest number of criteria. In order of effectiveness in prediction the three stanines were roughly pilot, bombardier and navigator. Among the individual tests the Rotary Pursuit and Two Hand Coordination Tests seem to have best predicted all of the criteria. Mechanical Principles and Reading Comprehension were approximately the next best in prediction.

Tests such as Numerical Operations, Finger Dexterity, Aiming Stress

TABLE 2.70.—*Summary of significant (1 percent or 5 percent level) correlations between predictor variables and criterion scores*

P–38 PILOTS—FOURTH AIR FORCE

Predictor variables	Criterion variables[1]							
	1	2	3	4	5	6	7	8
Bombardier stanine	**	**				**		**
Navigator stanine	*	*				*		**
Pilot stanine	**	**	*		*	**	**	*
Rotary pursuit CM803A	*		*		**	**	*	*
Two hand coordination CM101A	*	**		*		*	*	*
Complex coordination CM701A	**		*					
Aiming stress CM211A								
Discrimination reaction time CP611D	**							*
Finger dexterity CM116A								
Reading comprehension C1614G		*				*	*	**
Mechanical principles C1903A		**	*		*			*
Mechanical information C1905A								
Dial and table reading CP622-21A	**							
Spatial orientation II CP503B	*							*
Spatial orientation I CP501B								**
Numerical operations (F) C1702B								
Numerical operations (B) C1702B								*
Mathematics A C1702E		*						
Mathematics A C1702F								
Mathematics B C1710A								
Mathematics B C1206C								
Biographical data (P) CE602D							*	
Biographical data (N) CE602D								
Speed of identification CP610A	**							*
General information (N) CE505D								
General information (P) CE505D					*	*		
Technical vocabulary (Z) CE505C					*			*
Technical vocabulary (B) CE505C								*
Technical vocabulary (N) CE505C								
Officer quality					**	**		([2])

[1] Criterion variables are as follows:
1. Air-to-Air Gunnery
2. Air-to-Ground Gunnery
3. Skip Bombing
4. Dive Bombing
5. Mission Grades
6. Adversity Scores
7. Accidents
8. FEB Re-evaluation

[2] No data available.

and Vocabulary (N) and Biographical Data for Navigators had little predictive value for the criteria studied. Among the various criteria, skip and dive bombing were predicted very poorly by most of the tests and stanines.

Other Methods of Validation

Besides expressing degree of validity as a coefficient representing the correlation between test scores and a proficiency criterion, it would be possible to show relative validities by comparing the mean test scores of pilots of different levels of proficiency. The latter procedure has a great advantage in simplicity and ease of calculation, especially when the groups compared are at the extremes of the range of proficiency found. The tests which yielded the greatest and most consistent differences between the averages for the groups would naturally be those which best predicted degree of proficiency. By this method the numbers of individual pilots involved would be greatly reduced while at the same time little discriminating power would be lost since these groups may be expected to show to a maximum degree any relationship that existed.

For these reasons it was decided in the First Air Force to make the initial study of test validities by comparing the average test scores and stanines of two groups of pilots, one extremely high and the other extremely low in proficiency. Only for those tests which showed by this method a promising degree of validity would more extensive analyses be made and validity coefficients computed for the total group of pilots studied. It was not possible to complete this analysis in time for inclusion in this report

Evaluation

In the results presented above it would appear at first glance that at the level of operational training, classification test scores and the pilot stanine have a relatively small predictive value. However, in considering the validity coefficients obtained it should be remembered that there were a number of factors present which tended to put a "ceiling" on the size of the correlation coefficients to be expected between stanines and proficiency measures. Four factors might be mentioned here.

Restricted Range of the Pilot Stanine

Assuming that the unaugmented pilot stanines of original applicants for aircrew training were normally distributed from 1 to 9, data in table 23 show that less than 20 percent of the pilots reaching operational training had stanines as low as the median of the group of original applicants. The rejection at the classification centers of nearly all applicants with stanines of 1 and 2 and rejection of many with stanines of 3 and 4 resulted in the initial range restriction. The fact that in training in the AAF Training Command men with low stanines were much more frequently eliminated

than men with high stanines, produced a further limitation in stanine range in the group sent on to operational training.

Restricted Range of Proficiency

The elimination of most of the pilots of very low degree of proficiency in the AAF Training Command undoubtedly resulted in a smaller range of proficiencies in pilots entering operational training. Opinions of many training supervisors that all pilots in operational training were at least average and were satsfactory in proficiency tended to bear this out.

Inconsistency and Low Reliability of Criterion Scores

The reliability of a measure determines in part the maximum correlation that could be obtained with any other measure even if identical functions were involved in both cases. While the reliability of some of the criteria studied was as high as 0.70, it was in many cases much lower. And for several criteria it was not possible to obtain any estimate of reliability. In such cases the many known sources of uncontrolled variation would indicate a rather low reliability. Hence, many of the validity coefficients could only be low at best.

Dissimilarities in Primary Flying and Operational Training

Since tests and selection procedures were mainly validated against success or failure in primary training, they could not be expected to cover, except more or less by chance, any abilities needed in operational training which were not prominent in earlier training. Here a thorough validation of individual test scores against criteria of proficiency in operational training might lead to a different weighting of tests and specific abilities so as better to predict success at this and the combat levels.

The data so far obtained indicate that the pilot stanine best predicts comments on missions, ratings obtained by interviewing instructors, and time taken to complete all or part of training. Because of the relatively low coefficient of reliability of gunnery scores, it is also likely that the stanine predicts even better pilots' gunnery *ability* although the prediction of the gunnery *scores* is less accurate than the prediction of the other types of scores just mentioned.

SUMMARY

Research personnel in the First, Second, and Fourth Air Forces amassed a great amount of data on fighter-pilot proficiency in the P–38, P–47, and P–61 aircraft. The several types of potential criterion measures which were systematically explored and evaluated included scores obtained from air-to-air and air-to-ground fixed gunnery, dive and skip bombing, and rocket firing; comments and ratings from mission cards; tower and mobile-control-unit reports; ratings by instructors (on printed forms and obtained from

interviews); ratings and rankings by fellow students; amount of time required to complete training; involvement in pilot-error accidents; selection as instructor; re-evaluation by a Flying Evaluation Board.

Insofar as possible, the above potential measures of pilot proficiency were re-evaluated in terms of six criteria: distribution, reliability, number of men affected, discrimination between experienced ("expert") and inexperienced ("novice") pilots, completeness of coverage of the pilots' duties, and degree of objectivity.

It was concluded that the most satisfactory pilot proficiency measures were the objective scores derived from gunnery, bombing, and rocket firing; the specific comments entered on mission cards; and ratings (based on concrete reasons) obtained from interviewing instructors. It was further concluded that no single one of the measures adequately covered the full range of fighter-pilot duties. It was also felt that little dependability could be placed on any of the measures except where original preparation of the data was closely supervised at the training stations by aviation psychologists.

The pilot stanine was found best for predicting comments on mission cards, ratings obtained from interviewing instructors, time required to complete all or part of training, and re-evaluation by Flying Evaluation Boards. Stanine validity coefficients with the various criterion measures were lowered by such factors as the small number of cases (affecting particularly re-evaluated men, accident-pilots, and trainees selected as instructors), restricted range of both the pilot stanine and pilot proficiency at the operational level, and the low reliability of the criterion measures. This last factor was considered particularly to lower the degree of correlation between stanine and fixed-gunnery scores.

CHAPTER THREE

Photo-Reconnaissance Pilot

ANALYSIS OF DUTIES

Introduction

Most of the training of individuals and crews in the specialized field of photographic reconnaissance was done in the Third Air Force. Studies in this field were therefore made one of the responsibilities of aviation psychology personnel in this Air Force. Training in Photo-reconnaissance included four main types: F-3 training concerned with night photographic reconnaissance utilizing A-20 aircraft; F-5 training, target photo-reconnaissance and tri-metrogonic mapping (mosaics) in P-38 type aircraft; F-6 training, involving tactical reconnaissance, both visual and photographic, in P-51 aircraft; and F-7 training utilizing B-24 crews for long range precision mapping. Preliminary investigation suggested that the F-5 training program would be the most profitable area for initial research in photo-reconnaissance, since the available records were relatively well suited to validation studies. Investigation of the other three types was prevented by the curtailment of training at the close of hostilities. The discussions which follow in this chapter are therefore limited to the F-5 program.

Job Description

The F-5 training program consisted of four months training of P-38 pilots. It was divided into four phases: ground training, photographic transition in B-25 or F-10 type aircraft, training in P-38 type aircraft and, finally, night familiarization.

Ground Training

During this phase, instruction was given in operation of the aircraft, combat tactics, administration duties and general subjects. Training publications, films and information bulletins were used for instruction purposes. Elimination for failure in ground school subjects was very rare.

Photographic Transition

This phase was designed to acquaint the trainee with the fundamentals of navigation and aerial photography. The flights were in B-25 or F-10 type aircrafts. The trainee spent 5 hours as observer (standing behind the pilot and copilot), five hours as navigator and 5 hours as copilot.

Training in P-38 Type Aircraft

The largest portion of the four months training period was devoted to training in the F-5 type aircraft (P-38 modified for photographic reconnaissance purposes). It was necessary for the trainee to acquire complete knowledge of the performance data and operational limitations of the P-38. In addition, each trainee had to complete the following requirements:

1. A minimum of five photo-reconnaissance missions, with an average grade of 85 percent. At least one of the missions had to be above 30,000 feet. A minimum of two missions were required at the maximum tactical range of the aircraft.

2. One successful mapping mission at 20,000 feet or above, covering an area at least 10 by 20 miles in size. The grade for this mission had to be 75 percent or better to be considered successful.

Night Photo Familiarization

This phase of training was designed to acquaint the trainee with the capabilities and limitations of photo-flash bulbs, electrical flash equipment and proper techniques and tactics to be employed in the use of night cameras. The trainee was required to demonstrate competency in the operation of the K-19B camera and related equipment, and in the handling of bomb bays, shackles and release mechanisms. He was further required to know how to load, arm and inspect flash bulbs.

Job Specifications

No studies were made of special requirements for the job of the F-5 or photo-reconnaissance pilot, although some of the requirements may be inferred from the foregoing description of operational training. Other studies of higher priority prevented any detailed analysis of this type.

CRITERIA OF PROFICIENCY

General

Four types of grades were given in the F-5 program. The first consisted of the descriptive ratings "satisfactory," "very satisfactory," and "superior" on most ground school subjects. Several grades of "qualified" and "not qualified" were assigned in such subjects as code, blinker, altitude chamber and instrument flying. The third grade was "time in hours" for the various types of flying such as B-25 training, P-38 transition, and night flying. The fourth type consisted of numerical grades based upon a standard and more or less objective system of scoring short range photo, and long range photo and mapping missions. Preliminary investigation of these four systems indicated that the relatively objective mission or sortie scores were the best measure of proficiency. Main research effort was therefore concentrated on the sortie grade.

Sortie Grades

Description

The following procedure was used in assigning grades to photographic and mapping missions. The pilot was briefed on targets to be photographed, flying altitude, and flying time for the mission and course. Preflight of the plane and photographic equipment was the responsibility of the pilot. During the mission, the pilot was required to keep an accurate log with special reference to weather and flying altitude over each target. Upon return from the mission in addition to the photographic plates exposed, the pilot provided the following information to the Interrogation Office:

1. Name and number of targets assigned.
2. Name and number of targets claimed.
3. Mean photographic altitude.
4. Weather conditions over targets.
5. Camera lens size.
6. Description of targets of opportunity (if claimed).
7. Condition of equipment.

After the photographic plates were processed, the prints were sent to the Photo Laboratory for grading. The judgment as to whether a target was scored as covered, partially covered or not covered was dependent upon lens size as well as percent of target in the photograph. The scale used is presented in table 3.1.

After coverage was determined, the photo interpreter determined and recorded a sortie mission grade for the pilot listing the mission number, total time in the air, mean altitude of photography, number of targets assigned, number claimed, number of targets of opportunity claimed, number of targets not claimed with the reason given, and finally, for each size of lens used, the number of targets covered, partially covered and not covered. In obtaining the final grade, one credit was given for all targets covered, one-half credit for targets partially covered and no credit for targets not covered. The final mission grade, indicated in percent was the ratio of targets covered (total credit) to targets assigned plus targets of opportunity less nonpilot failures.

TABLE 3.1.—*Relation between percent of coverage and score assigned target photographs*

F-5 PHOTO RECONNAISSANCE—THIRD AIR FORCE

Lens size (inches)	*Percent coverage*	*Judgment or score*
6	0– 4	Not Covered.
6	5– 95	Partially Covered.
6	96–100	Covered.
12 by 24	0– 4	Not Covered.
12 by 24	5– 75	Partially Covered.
12 by 24	76–100	Covered.

In general, a single mission included from 12 to 18 assigned targets. To be regarded as complete, the mission had to cover at least 3 targets. Targets of opportunity could be photographed only in lieu of assigned targets which could not be photographed for legitimate reasons, or in addition to regularly assigned targets. Scores on both successful and unsuccessful missions, i.e. mission grade below as well as above 75 percent were used. However, the percentage of missions graded less than 75 percent was very small, less than 1.5 percent for a sample of 786 missions.

The coefficient of correlation between the average sortie grade for odd and even missions was found to be 0.62 for 152 pilots trained at Coffeyville Army Air Base during May and June of 1945. Other estimates of the reliability of these scores are presented in table 3.2.

TABLE 3.2.—*Reliability of photo-reconnaissance sortie scores*

COFFEYVILLE AAB—THIRD AIR FORCE

	Number cases	*r*
First short mission v. second short mission	157	0.27
First long mission v. second long mission	159	.32
Estimated reliability (corrected by Spearman-Brown formula)	[a]5	.63

[a] Missions.

Evaluation

Since the sortie grade was the only criterion studied, no comparisons are possible with other criteria. However, the following points can be made on the basis of the evidence available. First, it is not possible to determine the extent to which sortie grades discriminate between pilots known to differ in proficiency and experience since no such studies could be carried out. Second, scored sortie grades have a rather narrow dispersion. Whether this is due to the nature of the task in relation to method of scoring, or to actual similarity in level of proficiency of pilots in this type of operational training is not known. Third, the reliability of sortie grades, while not as high as might be desired, is sufficient for them to be useful in validation studies. Fourth, in terms of coverage of the job and logical validity these grades are an excellent criterion. Finally, since the pilots concerned are not known to those assessing target photographs, sortie grades are not subject to most types of subjective errors. Estimates of percent of targets appearing in the photographs, and decision as to whether a failure to attempt photographs occurred as a result of causes beyond control of the pilots, are possible sources of errors of various sorts. In general, it may be concluded that the sortie grade should provide a useful criterion of proficiency in operational training.

VALIDATION

Average sortie grades of trainees at the two Third Air Force stations training F-5 pilots were correlated with their aptitude scores or stanines. In this comparison, only trainees who had completed a minimum of 6

TABLE 3.3.—*Correlations between average sortie grades and stanines*

F-3 PILOTS—THIRD AIR FORCE

Station	Number cases	Sortie Grade		Bombardier			Navigator			Pilot			Pilot and credit		
		mean	S. D.	Mean	S. D.	*r*	Mean	S. D.	*r*	mean	S. D.	*r*	Mean	S. D.	*r*
Coffeyville	349	93.0	5.6	6.00	1.77	**0.18	5.50	1.73	**0.23	6.09	1.49	.08	6.24	1.57	*0.10
Will Rogers	121	87.1	6.0	6.01	1.71	*.23	5.52	1.88	.16	6.41	1.35	.06	6.70	1.64	.07
Total[1]	470	......	......	......	......	**.20	......	......	**.23	......	......	.07	......	......	*.09

[1] Average by Fisher's z.

photographic missions were included. The coefficients of validity obtained are shown in table 3.3 together with the appropriate means and standard deviations.

Attention is called to the relatively higher validity of the navigator stanine. This result should be interpreted with caution, however, since the group had been restricted with respect to the pilot stanine. Validity coefficients for both the bombardier and navigator stanine are significant at the 1 percent level and were regarded as sufficiently high to warrant the computation of test validities. This was not done, however, due to the shortage of time.

Separate grades were available for short and long missions for about one hundred trainees at Will Rogers Field. Comparison was therefore made between stanines and grades on each type of mission to find out whether length of mission was a factor in stanine validity. Validity coefficients with unaugumented pilot stanine were 0.11 and 0.16 respectively for short and long mission (sortie) grades of 105 F–5 pilots for whom separate scores were available. Although the difference between the coefficients is not statistically significant, there is a suggestion that the pilot stanine better predicts the grade on long than on short missions.

SUMMARY

Research by aviation psychologists in the area of photo-reconnaissance training was limited to the F–5 program. This program was concerned with the training of P–38 pilots in target photo-reconnaissance and trimetrogonic mapping (mosaics). One criterion measure for this type of training was investigated, that of objective mission grades for photographic reconnaissance missions. The grades were based on the extent to which assigned targets were covered in the photographs obtained by the trainee. The reliability of the mission grade criterion was found to be satisfactory for validation purposes. For two samples of F–5 trainees for whom both stanines and mission grades were available, stanine validities were, in general, low; however, several coefficients were higher than could be expected by chance alone.

CHAPTER FOUR

Airplane Commander

ANALYSIS OF DUTIES

Introduction

It was the mission of the Continental Air Forces to train in combat type operations combat crews made up of individuals who had completed training in their individual specialties. Because of the primary emphasis upon the training of crews, training personnel paid relatively little attention to individual proficiency, most of the attention being given to crew performance. Although aviation psychologists in the Continental Air Forces were similarly interested mainly in crew proficiency, two factors led to the expenditure of much time and energy in obtaining evidence of individual proficiency as well. The first of these was a desire to obtain data for use in studying the validity of individual selection and classification procedures. The second was the fact that since no satisfactory objective measure of it was available, crew proficiency was thought by many to be best observed in the individual proficiency of various crew members, particularly the airplane commander.

The airplane commander was probably the most important member of the combat crew. As commander of his crew, he was responsible for their performance as well as his own. From the point of view of psychological research in operational training, the airplane commander was also important because he was responsible for much of the training of the rest of the crew. His many contacts with instructors and training personnel made him by far the best known member of the crew. It was thus expected that more and better data would be available on the proficiency of the airplane commander than would be available for any other aircrew specialty. And, as will be brought out later, the proficiency of the crew as a whole and that of the airplane commander were usually considered nearly identical.

Job Description

Because of his special importance in training and in combat operations, research personnel in the Continental Air Forces studied the job of the airplane commander in operational training in greater detail than they did the jobs of the other crew members. Information was obtained from a variety of sources. AAF Training Standards 20-2 and 20-3, Training Manuals such as Second Air Force Manuals 50-27, 50-51, 50-126 and similar Manuals

of the Third and Fourth Air Forces were consulted. Interviews were held with Directors of Training, with instructors, with airplane commander trainees and with officers returned from combat in an attempt to obtain as much information as possible about the duties and requirements of the job of airplane commander.

The usual heavy and very heavy bombardment crew training in the Continental Air Forces covered a period of 12 weeks. It included both flying and ground school training. A minimum of 120 hours of flying was required of each crew, although in practice most crews needed more flying hours than this to complete the required operations. Flight training was set up in the form of specific missions of many types, with the number of hours in the air needed for each varying according to the mission requirements. Included were relatively short and long range missions, missions at high and low altitudes, night and day missions, bombing, navigation and fighter interception missions. A total of 26 different missions were required in B–29 (VH) training, several of which had to be repeated to meet minimum requirements. In B–17 and B–24 (H) training, a larger number of missions was required; but the length of time needed to accomplish the individual missions was correspondingly less. Details of each mission were specified in flight training directives which listed the operatons to be performed and the responsibilities of the various crew members.

Similarly, the ground training of airplane commanders and other crew members was specified in Technical Training Manuals and ground training directives. Very little of the ground training was given to crews as units, separate classes being given for individuals in each specialty. Airplane commanders took courses in the technical aspects of plane operation, communications, navigation and bombing. Courses were required of most crew members in intelligence, survival and the usual military subjects. The airplane commander and certain of the other crew members took courses in combat tactics. In addition, the airplane commanders took 6 hours of ground school under the general headings of duties and responsibilities of the airplane commander. Ordinarily there were no particular objectives or requirements for these 6 hours and individual instructors utilized them as deemed desirable. Sometimes the time was devoted to conferences on the problems which had arisen in the handling of the crews and related matters. At other times training films were shown and lectures given by returned pilots.

The responsibilities of the airplane commander as leader and commander of the crew were usually given extra emphasis in some way. At orientation lectures, informal conferences with instructors and in the appropriate ground school courses suggestions were given as to what these responsibilities were and how they could be met. These included suggested means of utilizing "waiting hours," e.g. time spent waiting for transportation, briefing, weather clearance, etc.; ways of gaining knowledge of crews' background; methods of instruction in crew specialties; and so on. Flight train-

ing records were usually made available to the airplane commander. On check rides, notes were sometimes made on his effectiveness as commander. However, the latter were informal and haphazard and no systematic records or checks were kept on the ability of the pilot to command his crew.

At many stations, the airplane commander was given certain definite responsibilities for the training of the rest of the crew. These included responsibility for the punctual attendance of each crew member at his classes and formations. Absences, tardiness, etc., were reported to the airplane commander who had to present to training officials the reasons therefor and disciplinary action taken to prevent recurrence. Deficiencies in class work and in flight were reported to the airplane commander for action to correct them. He was often responsible for keeping the crew informed of training progress. In addition to the regular briefings, the airplane commander was encouraged to give additional explanations to his crew specifying the part each crew member was to play in accomplishment of the mission. Review of previous mistakes and suggestions for improvement were to be given. At the conclusions of missions, the airplane commander was frequently required to hold a critique of the mission with the crew, discussing the performance of each crew member. The airplane commander was responsible for seeing to it that crew members kept busy on all missions, assisting other crew members if their own specialty was not needed.

Probably the best summary of the duties and responsibilities of the airplane commander in operational training is that given in AAF Training Standard 20-2:

> The pilot will be trained to a standard enabling him to handle his aircraft with skill and confidence. In addition to piloting skill, he must be fully instructed in his duties and responsibilities as airplane commander to insure that he exercises proper supervision over the personnel and equipment under his control. Preflight inspection should be systematically made by crew members and reports rendered to the pilot. Prior to every training mission, the pilot will insure that his aircraft has been thoroughly checked and inspected to include complete coverage of current check list. During every training mission, the duty of the pilot will be to supervise the functioning of each crew member with a view to developing individual skill and building his crew into an efficient combat team.

Although all of the flying and technical ground training of airplane commanders was specified in considerable detail, it was frequently not possible to maintain uniformity of training at the station level. And, while some attempt at least was made to standardize flight procedures, missions and ground school courses, there was very little attempt to standardize the procedures to be used in training airplane commanders in the exercise of command. Thus training in accomplishing the functions of crew commander varied even more than training in pilot skills. These variations added greatly to the difficulties met in carrying out research on the proficiency of airplane commanders. A further discussion of these difficulties and of the factors contributing to the lack of uniformity in training conditions will be given in the sections of this chapter dealing with criteria of proficiency.

Airplane Commander Check List

While flying skill was, of course, of equal or even greater importance in operational training, psychological research on the job of the airplane commander emphasized his activities as commander of the crew rather than the former. This emphasis had its origin in the belief that the former aspect of the job could best be studied in connection with training of the individual pilot in the AAF Training Command. Command of the crew was felt to be the chief new or unique feature of the activities of the airplane commander in operational training. To obtain more detailed information on the job of crew commander, research personnel in the Third Air Force made a series of direct observations of crew training. An aviation psychologist was attached to a class of combat crews at Avon Park AAF to make a study of the activities of the airplane commander in operational training. This officer flew 18 of the 26 missions constituting the minimum requirements. Several of the missions were flown more than once; and flights were made with more than half of the crews in the class. Logs were maintained during all of the flights which included detailed observations of the activities of the airplane commander. Records were also kept of observations of the crew on the ground and of conversations with crew members. On the basis of this study, and utilizing the results of other analyses and suggestions, an Airplane Commander Check List was prepared in the Third Air Force. In this check list were included 27 items, each referring to a training situation in which it was believed degree of success in the exercise of command or a fundamental pilot skill could be observed. In using the check list, instructors or other observers flying with an airplane commander recorded his effectiveness in the specific situation referred to in the item. A copy of this check list is found in the appendix G.1. These items may be considered a sample of the more observable features of the job of airplane commander.

Job Specifications

Two types of approach were utilized by aviation psychologists in arriving at specifications of the characteristics needed by airplane commanders for success in operational training. The first approach was to collect and summarize the opinions of experts—i.e. instructors, combat returnees, flight surgeons and others. The second was to study the reasons for failure in pilots known to have failed in operational training as airplane commanders.

Characteristics Suggested by Experts

The detailed procedures whereby aviation psychologists obtained from instructors, combat returnees and others opinions as to characteristics needed for success in operational training varied somewhat in the different air forces. The relative emphasis upon different aspects of the job also varied in these interviews, probably in part depending upon current research emphasis. For example, in the Second Air Force, approximately 75 officers with combat experience in lead crew or command position were

interviewed in an attempt to discover what were the most important characteristics needed for success in lead crew position. While the interviews were thus specifically oriented toward crew characteristics, personnel interviewed tended without exception to describe what they considered the most important traits or characteristics of various crew members. Thus as a by-product of interviews designed to obtain characteristics of lead crews there was obtained a series of descriptions of what were thought the most important characteristics of successful airplane commanders. In the Third and Fourth Air Forces similar interviews had an entirely different emphasis since they attempted to obtain direct descriptions of successful airplane commanders. In addition to such interviews, in the Third Air Force the research personnel collected descriptions of situations in which some airplane commander had exhibited either good or poor proficiency in combat. In all, 565 descriptions of such combat situations were obtained. Also obtained were about the same number of descriptions of similar situations in operational training. These situations were sorted into categories by the research personnel according to trait or characteristic exhibited by the airplane commander described.

From interviews with experts, from descriptions of situations, from study of Training Manuals and Standards and from observations of training, the following traits or characteristics seemed to research personnel in all three air forces the most important for success as airplane commander. Included are items referring both to skills and to aptitude, personality and character. The listing is by no means exhaustive; nor are items necessarily listed in order of importance.

1. *Skills.* a. *Ability to fly well in formation.*—Formation flying was felt by most experts to be the most important feature of bombardment operations in combat. Aspects of formation flying that were especially stressed were: smoothness of control, so that position in formation was maintained continuously; knowledge of correct procedure in assembly, in formation turns and evasive action, etc.; and foresight, or ability to plan the proper maneuvers well in advance.

b. *Ability to fly accurately on instruments.*—In heavy and very heavy bombardment flying, instruments played a much greater part in plane operation than in the case of fighter aircraft, particularly in formation flying.

c. *Knowledge of the aircraft.*—Thorough familiarity with the characteristics of the aircraft was necessary for accurate instrument and formation flying and for early detection and forestalling of incipient emergencies by prompt action.

d. *Ability to land and take off with skill.*—Superior pilot skill was often called for in unique landing situations. Since nearly half of the aircraft accidents in operational training occurred in handling aircraft on the ground, skill in landing and taking off was probably more important in operational training and combat than was ordinarily admitted.

2. *Traits or personal characteristics.* a. *Eagerness or enthusiasm.*—Observations in the Continental Air Forces confirmed the belief that here, as well as elsewhere, motivation was one of the most important factors in job success. Eagerness and enthusiasm accentuated other favorable traits and made more effective whatever skills the airplane commander possessed.

b. *Good judgment.*—In interviews with experts the term "judgment" was suggested in many different contexts. In general, the term referred to the making of successful reactions on the basis of rapid evaluation of many different aspects of the immediate situation. The situations suggested varied from those in which the important features were the spatial relations between planes or between plane and the ground to extremely complex situations in which these and other factors such as degree of damage, fuel consumption, crew injuries and the like were considered in arriving at a decision whether to abort a mission, ditch the aircraft or take other appropriate action.

c. *Interest in other crew members.*—Since the airplane commander was responsible in part for training his crew, the degree of interest he showed in them was reflected in the effectiveness of their training. Motivation of the other crew members and crew morale were, of course, directly influenced by the interest of the airplane commander in them.

d. *Leadership.*—As was the case with "judgment" the term "leadership" was used in a variety of contexts. In general, it referred to the over-all effectiveness with which the airplane commander handled the crew. Three aspects of leadership received special emphasis. The first was willingness to assume responsibility. The attitude of "Let George do it" was rarely found in a successful airplane commander. A second aspect was the ability to cooperate with other crew members in joint activities. Proper assignment of responsibilities and confidence in and proper evaluation of the contributions of other crew members were most important. Finally, a general over-all likeableness was recognized as important in reducing irritations between crew members and in keeping up crew morale. This also resulted in smooth relationships between crew and training personnel since the airplane commander represented the crew in most of these contacts.

e. *Conscientiousness.*—Absolute accuracy in carrying out preflight inspections, exact following of SOP and following without change the exact briefed instructions were considered important in maintaining flying safety and in successful accomplishment of combat missions.

f. *Calmness in emergencies.*—Ability to deal calmly and matter-of-factly with emergency situations resulted in correct decisions being made by the airplane commander himself and in the maintenance of good morale and effective crew discipline under stress of combat.

g. *Foresight.*—The ability to plan ahead and foresee the outcomes of present situations enabled airplane commanders to guard against emergencies. Crew members could be warned to take appropriate precautions and smooth teamwork achieved even in relatively sudden emergencies.

Reasons for Flying Evaluation Board Action

The reports of Flying Evaluation Board action were analyzed to discover if possible the main reasons for failure in operational training. Data were available for 33 bombardment pilots in the Third Air Force and 14 bombardment pilots in the Fourth Air Force who were re-evaluated in 1944. In these analyses an attempt was made to determine from each Board Report the principal reasons for reclassification of the pilot concerned. In table 4.1 are given the frequencies with which various categories of reasons for re-evaluation occurred. The great variation in material included in these reports makes impossible any more detailed or meaningful analysis of the reasons for failure.

TABLE 4.1.—*Reasons for reclassification of pilots by FEB board action*
PILOTS RECLASSIFIED IN 1944 IN THIRD AND FOURTH AIR FORCES

Reason for reclassification	Number of cases	
	3AF	4AF
Physical disability	0	3
Lack of proficiency	2	4
Defects in personality	31	12
Fear	......	(7)
Personal misconduct	......	(3)
Lack of maturity	......	(2)
TOTAL	33	19

Relative Importance of Various Skills and Traits

Two methods were utilized in arriving at an estimate of the relative importance of various skills and characteristics for success in the job of airplane commander. One was through the frequency with which the trait or characteristic was mentioned by experts and the other was through ranking of the traits or characteristics by instructors and others.

1. *Frequency of mention in situation descriptions.*—The relative frequency with which various skills and traits were mentioned in interviews with combat returnees, instructors, etc., was one way of estimating their degree of importance in the job of airplane commander. This method is open to the objection that frequency of mention in interviews may be dependent more upon frequency of occurrence or ease of observing these skills or characteristics than upon any fundamental importance in combat or operational training. However, as a matter of general interest, the frequency with which various traits and skills were mentioned was used in the Second and Third Air Forces for estimating their relative importance. Since the Second Air Force interviews were directed primarily toward obtaining information about lead crew proficiency, frequencies of mention of airplane commander traits in these interviews was probably biased by that emphasis and such data were not included in the discussion which follows.

Research personnel in the Third Air Force obtained more than 500 descriptions each of combat and training situations illustrating good or poor airplane commander behavior. These situations were analyzed to determine

the traits or characteristics referred to in each situation. The frequency with which various traits and skills appear in these descriptions is of some interest, particularly the comparative frequency of mention for combat and training operations.

Half of both the training and combat situation descriptions referred to skill in the handling of emergencies and unusual situations. When these were sub-divided as to type, certain differences between combat and training appeared. Formation flying and handling of emergencies in bombing and strafing of targets appeared more ofter in combat situation descriptions, while handling of mechanical failure or fire appeared more often in the training situations. In general, skill in routine flying was described more often in the training situations. Here landing and preflight checks were the most frequently mentioned activities and were also mentioned with greater relative frequency in training than in combat situations. Strict adherence to SOP, maintenance of authority, and air and ground discipline were referred to much more often in the descriptions of combat situations than in those of training situations. Proficiency of the crew as a whole and miscellaneous other characteristics or activities were mentioned infrequently and occurred about equally often in combat and training situation descriptions.

2. *Trait rating by instructors.*—In the Fourth Air Force research personnel asked 7 or 8 instructors at each of three stations to rank 10 traits or skills in the order of their importance for success in operational training. Table 4.2 lists the 10 traits ranked in order of the combined ranking by the 22 instructors from the three stations, together with the average rank and standard deviation of the ranks for each item. The rank-difference correlation coefficient between the average ranks given by the odd and even numbered instructors was 0.99, showing there was almost perfect agreement among groups of instructors as to relative trait importance.

In addition to ranking the 10 traits or skills each instructor at two of the stations rated his airplane commander trainees on each of the 10 traits. Ratings were given on the basis of upper and lower half designations. The instructors also gave a general rating of over-all proficiency. After the ratings had been obtained, tetrachoric coefficients of correlation were computed to show the degree of relationship between the rating made on each separate trait and the over-all or total rating. In this manner it was hoped to obtain additional evidence as to the relative importance of each trait or skill with respect to total proficiency. The 10 traits were then compared in rank as given by the instructors when they ranked the traits in order of importance and in the degree of correlation between ratings of airplane commanders on the traits and ratings of total proficiency. This comparison is also given in table 4.2. Certain interesting differences are seen between instructors' beliefs as to trait importance and the relative weight given traits in arriving at a rating of over-all proficiency. The most striking difference is in the trait of "likeableness." Instructors considered it the trait of least importance when they ranked the traits in order of importance.

However, when they rated the over-all proficiency of the airplane commander trainees, apparently likeableness played a larger part than any other trait. Similarly, "landing skill" was considered eighth in importance when the traits were ranked but was third in importance when they rated their students. The opposite type of relationship occurred with "concern with safety of crew." This was thought third in importance when the instructors ranked the traits but was eighth in importance in their ratings of students.

TABLE 4.2.—*Comparison of trait rankings with size of correlation between trait ratings and over-all ratings*

AAF MUROC AND WALLA WALLA—FOURTH AIR FORCE

Trait or skill	Average rank[1]	S. D.	Order of importance by ranking[1]	Order of importance by size of r[2]
Knowledge of the airplane	1.77	1.1	1	4
Performance in an emergency situation	3.18	1.9	2	2
Concern with safety of airplane and crew	3.77	1.4	3	8
Eagerness	4.46	2.6	4	5
Adherence to briefing instructions	5.09	2.0	5	7
Skill in formation flying	5.23	2.3	6	6
Ability to get along with crew	6.36	2.2	7	9
Skill in landings	8.09	1.3	8	3
Skill in take-offs	8.18	1.2	9	10
Likeableness	8.86	1.6	10	1

[1] The 10 traits were ranked by 22 instructors at the above stations and at March Field.

[2] The correlations between trait and over-all ratings were based on ratings of 86 airplane commanders at the above listed stations.

CRITERIA OF PROFICIENCY

Introduction

As with fighter pilots in chapter 2, in this chapter on the airplane commander, and in the following chapters on other aircrew specialties, there is separate discussion of each major criterion of proficiency studied. An attempt is made in each case to evaluate the criterion in terms of its usefulness in validation studies and other research. In the discussion which follows, it will be shown that although the airplane commander was the most important and probably the best known crew member, it was very difficult to obtain adequate criteria of his degree of proficiency. Thus most of of the studies that are reported dealt with various types of ratings and other subjective estimates of proficiency. Few satisfactory criteria of an objective type were found to be available.

Ratings of Proficiency

Over-all Judgments

1. *5-point ratings.*—a. An obvious source of over-all judgments of proficiency of Army officers was the routine *efficiency rating* on WD AGO Form 66-2, 66-3. Although knowledge of the way such ratings were frequently made led to little expectation that they would be useful, their availability and the absence of other measures led to their use in exploratory studies.

These efficiency ratings were made on the following 5-point scale: unsatisfactory, satisfactory, very satisfactory, excellent and superior. Table 43 presents a typical distribution of these ratings.

TABLE 43.—*Distributions of 5-point ratings of airplane commanders*

SECOND AND FOURTH AIR FORCES

Efficiency Ratings, Form 66-2 (Fourth Air Force)

Rating:	*Number of cases*
Superior	10
Excellent	38
Very satisfactory	4
Satisfactory	0
Unsatisfactory	0
Total number	52

Over-all Descriptive Ratings (Second Air Force)

Rating:	*Number, Scale A*	*Number, Scale B*
1	43	56
2	136	75
3	114	98
4	20	53
5	7	16
Total number	320	332

Very little evidence is available regarding the reliability of the efficiency ratings. For one class of 52 airplane commanders at AAF Ardmore in the Second Air Force, 2 efficiency ratings were made approximately 30 days apart. While the first set of ratings was not consulted when the second set was made, the same administrative personnel were involved in both. Hence the correlation of 0.48 obtained between the two sets of ratings does not represent the reliability of truly independent efficiency ratings. No other data on reliability were available.

b. Very similar to the efficiency ratings on Form 66-2 were general ratings of *officer quality* and *technical* skill made at another Second Air Force Station on 55 airplane commanders in one class. The same 5-point scale was used as for the efficiency ratings. The distributions of scores obtained were also similar to those obtained with routine efficiency ratings. No direct evidence was available as to the degree of reliability.

c. As part of a project utilizing descriptive ratings in the study *of lead crew proficiency*, over-all 5-point ratings were made of airplane commanders by instructors in the Second Air Force. A description of the scales used and the procedures of administration is given in the discussion of descriptive ratings which follows in a later section of this chapter. Two scales, A and B, were used and involved somewhat different 5-point ratings. In Scale A, airplane commanders were rated on over-all proficiency in the same categories as were used in routine efficiency ratings. In Scale B, the following categories of proficiency were used: "This airplane commander is in the"—highest 10 percent, upper 20 percent, middle 40 percent, lower 20

percent and lowest 10 percent. Distributions of ratings obtained with both scales are given in table 4.3 above.

At a total of 4 stations in the Second Air Force, 2 or more instructors rated each of approximately 150 airplane commanders with Scale A, and similar ratings were made at 4 more stations on 160 airplane commanders with Scale B. The correlations between ratings by different instructors were computed separately for each station and combined by the use of Fisher's z into a single reliability coefficient for each scale. The over-all ratings of airplane commanders showed reliabilities of 0.39 and 0.66 for Scales A and B respectively.

2. *Miscellaneous ratings.*—a. Other ratings made routinely in the Continental Air Forces during 1944 were the ratings on the *Officers Proficiency Card*, AFTRC Form 2, a training form accompanying each officer from the AAF Training Command to operational training. At the end of the first month of operational training, ratings were entered on this form together with a statement regarding any deficiencies in previous training that might have been noted. The form was then returned to the AAF Training Command where it served as a sort of validation of training and as a guide for training revision. In practice the handling of these forms was most irregular. They were frequently lost, ratings were not entered properly, and the forms rarely served any useful purpose.

Three general ratings were made on the Officers Proficiency Card, one each under "military discipline," "manner of performing duties," and "general aircrew ability." Each rating was on a 3-point scale of below average, average or above average. Ratings of airplane commanders of 3 classes at AAF Ardmore were obtained on this form.

No direct evidence of degree of reliability was available.

b. At Biggs Field in the Second Air Force, *flight surgeons rated* each airplane commander in one combat crew class in terms of over-all proficiency. A 3-point scale of below average, average and above average was used. No data bearing directly on reliability were obtained and the flight surgeons reported that their contacts with combat crews did not give them sufficient information to make adequate ratings of this type.

3. *Rankings.*—Another procedure for obtaining over-all ratings of individuals is to have the individuals placed in an order-of-merit series. Instructors at several Second Air Force stations were asked to list the airplane commanders they knew in their current classes in the order of the expected effectiveness of these airplane commanders in combat. The organization of training was such that in practice each instructor usually knew only the airplane commanders in a single flight. The flights varied in size from one station to another, having from 9 to 28 crews. At one or two stations an instructor was found who felt he knew all airplane commanders in a class. In these cases the whole class was ranked in a single list.

Since size of flight varied, all rankings were converted to equivalent standard scores for statistical studies. Distributions of scores were thus

"normal" and all flights and classes had equivalent scores regardless of possible differences between them in a level of proficiency. At 3 of the stations in the Second Air Force, two or more instructors ranked the airplane commanders in a class more or less independently. That is to say, the actual rankings were done independently, although the resultant rankings may not have been truly independent, for there was no control over the extent to which the two instructors may have previously talked over the relative proficiencies of the various airplane commanders rated. At two of the three stations, the independent rankings were by flights with correlations of 0.81 and 0.45 respectively between rankings by different instructors. At the third station, the whole class was ranked together and the correlation between rankings by two instructors was 0.38.

Descriptive Ratings.

1. *Upper and lower-half ratings of traits and skills.*—From an analysis of observations during training, of interviews with training officers and from training manuals and the like, aviation psychologists in the Fourth Air Force obtained a list of 10 traits or characteristics of airplane commanders thought the most important in operational training. A rating scale was then constructed for those items by stating each trait or characteristic in the form of a question. The instructors were to rate each trainee as either above or below average with respect to each item, placing half of the trainees in each category. The list of traits and characteristics has already been presented in table 4.2 above.

Instructors at 3 stations in the Fourth Air Force were asked to rate the airplane commanders whom they had checked on training flights in terms of each of the traits and also in terms of over-all proficiency. In order to minimize "halo" effects, each item or question was placed on a separate sheet of paper together with the names of all students in the class or section. In this manner ratings were obtained from 2 or more instructors for a total of 55 airplane commanders at 2 B-24 Stations. The reliability of these ratings was estimated for each of the items by comparing the ratings given the same airplane commanders by different instructors. Separate coefficients were computed for each station and the coefficients combined by Fisher's z to give a single coefficient of reliability.

Of the various item ratings, those of "over-all proficiency," "likeableness" and "eagerness" had a fairly high degree of reliability, the coefficients of correlation between ratings of different instructors being 0.84, 0.70 and 0.79 respectively. (Since the ratings were in two categories only, all coefficients of reliability were based on tetrachoric coefficients of correlation). Ratings on the remaining items had reliability coefficients between 0.41 and 0.61. It is interesting that as with other scales in other air forces, there was better agreement among raters when they were rating over-all proficiency and personal traits than when they were rating specific skills and behavior. The crudeness of these ratings (into two categories only) and the

presence of the usual subjective errors make them somewhat less useful than would be indicated by the reliability coefficients obtained.

2. *Five-point ratings of lead crew proficiency.*—AAF Letter 50-117, 7 June 1945, directed that aviation psychologists in the various Continental Air Forces cooperate with Operations and Training personnel in the development of procedures for designating some of the combat crews in each CCTS class as potential lead crews at the end of their operational training. Since no readily available objective measure of crew proficiency had as yet been found and since it appeared unlikely that one would be found in the near future, it was decided in conferences with the research personnel concerned that three main types of rating devices would be developed and tried out, one in each of three air forces. The three types of devices and the air force of primary responsibility were: descriptive ratings of lead crew proficiency, Second Air Force; check lists of behavior and traits important in lead crews, Third Air Force; and ratings of mission performance at a lead crew school, Fourth Air Force.

In the development of descriptive ratings of lead crew proficiency in the Second Air Force, the usual preliminary steps were taken. Approximately 75 officers and enlisted men with combat experience in lead crew or command positions were interviewed to obtain their opinions as to what were the most important traits or characteristics of crews for success in lead crew position. From these interviews were obtained descriptions of a few characteristics applicable to the crew as a whole. For the most part, however, those interviewed described traits and characterstics of individual crew members. It was decided, therefore, that in addition to descriptive ratings of the crew as a whole, it would be desirable to utilize descriptive ratings of individual crew members in those characteristics considered most important for lead crew success. On the basis of the interviews and from study of lead crew training manuals, combat intelligence reports and other sources, a list of the most important crew and individual traits and characteristics was prepared. Certain traits felt by research personnel to be non-ratable for some reason were discarded and rating scales were then constructed for the rating of individuals and crews on the remaining traits and characteristics.

A number of considerations determined in part the form of the rating scales decided upon. On the basis of experience with descriptive ratings in civilian situations, it was felt that in order to obtain optimum discrimination between individuals and crews it would be desirable to have ratings on from 7 to 11 step intervals for each characteristic rated. However, previous analysis of routine proficiency ratings by instructors and supervisory personnel had indicated that they ordinarily utilized no more than three out of five available steps in making such ratings. Therefore, 5-point or 5 step-interval scales were decided upon with the thought that this would be a practical compromise between the usual three and the ideal seven rating steps. It was believed that with improved instructions and greater supervi-

sion of ratings all 5 step-intervals would be used by the raters. As with most descriptive rating scales, not only was the trait or characteristic to be rated described, but each step or rating point was described with a brief statement representing the typical individual to be rated at that point.

After the decisions had been made as to the main traits and characteristics to be included, two separate and distinct rating scales, Scale A and Scale B, were developed in the 16th and 17th Bombardment Operational Training Wings of the Second Air Force. Two scales were developed (rather than a single one for use throughout the Second Air Force) largely for reasons of administrative convenience. Because of the need for the immediate trial of some such device, there was not time for the research personnel in the two wings to get together on a single scale. Each group, therefore, developed a scale of its own within the framework of the general instructions issued. Thus Scale A and Scale B differed considerably in certain features of the scales. In Scale A the descriptions of the different step intervals for the rating of a single trait or characteristic were quite detailed and, more important, were almost identical in wording for all the step intervals. The difference between the step intervals or rating points was designated by expressions of quantity or frequency such as "always," "sometimes," "most" and "average." In Scale B, the descriptions of the steps for the ratings on an item tended to be more brief and to refer more to specific bits of behavior thought by research personnel to represent different degrees or amounts of the trait in question. Thus in Scale B the wording differed more from one step or point to another.

Scales A and B also differed somewhat in the traits and characteristics included for rating. Those included for the rating of the crew as a whole and for specialties other than the airplane commander are described later on in the appropriate chapters. For the airplane commander, Scale A included the following items: foresight, interest in his crew, proficiency in formation flying, proficiency in instrument flying and over-all lead crew proficiency. Scale B included items on proficiency in formation flying, eagerness and enthusiasm, foresight and planning, leadership, proficiency in instrument flying, and two general items, one on over-all lead crew proficiency and the other calling for a ranking of the airplane commanders in order of their proficiency.

With both scales mimeographed blanks containing the descriptive statements were prepared in advance. On these blanks were typed the names of the individuals and crews to be rated. Instructors or training officers were then asked to rate each individual or crew known to them. The rating step descriptions were listed in order from one to five with one representing greatest degree of proficiency. The printed directions for making the actual ratings were identical for the two scales. Instructions were also given that different instructors rating the same individuals or crews should not discuss their ratings with each other. An attempt was made with each scale to obtain at least two relatively independent ratings for each individual and crew rated.

There was one additional difference between the scales as they were actually administered. With Scale A the actual making of the ratings was not supervised by the research personnel. In this respect the ratings obtained were perhaps similar to what could have been obtained more or less routinely through channels, or by mailing the scale to the Directors of Training concerned. Spot checks made for that purpose showed that the instructions to avoid collaboration in making the ratings were followed quite well. With Scale B the research personnel attempted wherever possible to actually be present while the instructors were making their ratings. No record was kept, however, as to whether or not a particular set of ratings was directly supervised and in some cases no supervision was possible. With both scales instructions were given that no ratings were to be made for individuals and crews not known to the rater. Complete copies of Scales A and B may be found in appendix F.1 and F.2.

During the summer of 1945, Scales A and B were tried out with a number of B–29 crews in both CCTS and OTU training. In the 16th BOT Wing ratings were obtained with Scale A for a total of 160 crews at four combat crew training stations. In the 17th Wing ratings were secured with Scale B on between 150 and 200 crews, approximately 100 of which were in OTU training at 2 stations, the remainder being in CCTS training at 2 other stations. The distributions of the sums of the item ratings of airplane commanders obtained with the two scales are given in table 4.4. It can be seen that with both scales there was good dispersion of the totals of the item ratings.

TABLE 4.4.—*Distribution of airplane commander ratings*

SCALES A AND B, SECOND AIR FORCE

Scale A, 5 Items Rating:	*N*	*Scale B, 7 Items* Rating:	*N*
5	9	6–7	16
6	12	8–9	22
7	8	10–11	20
8	16	12–13	30
9	10	14–15	61
10	23	16–17	30
11	34	18–19	40
12	38	20–21	66
13	50	22–23	26
14	43	24–25	15
15	50	26–27	11
16	16	28–29	10
17	6	30–31	8
18	3	32–33	2
19	1		
20	1		
Total	320	Total	357
Mean	12.20	Mean	17.33
S.D.	2.97	S.D.	5.78

The reliabilities of the ratings given for each airplane commander trait or characteristic are shown in table 4.5. The coefficients given in the table

are the averages (by Fisher's z) of the correlations obtained at the different stations between ratings of the same airplane commanders by two different instructors. The variations in reliability from station to station are illustrated in the reliabilities of the total ratings on all items which are shown for each station separately in table 4.6.

TABLE 4.5.—*Reliability coefficients for airplane commander rating scale items*

16th AND 17th WINGS—SECOND AIR FORCE

Scale A

Item	N	First Rater, 1		Second Rater, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
1. Foresight	160	2.39	0.77	2.56	0.77	[illegible]
2. Interest	160	2.30	.65	2.35	.73	[illegible]
3. Formation	160	2.28	.71	2.43	.72	[illegible]
4. Instrument	160	2.63	.63	2.62	.65	[illegible]
5. Lead possibility	160	2.45	.94	2.38	.79	[illegible]

Scale B

Item	N	First Rater, 1		Second Rater, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
1. Lead possibility	151	2.67	1.14	2.65	1.12	[illegible]
2. Formation	115	2.47	.96	2.56	.91	[illegible]
3. Eagerness	133	2.77	.98	2.38	.86	[illegible]
4. Foresight	149	2.53	1.00	2.41	.85	[illegible]
5. Leadership	136	2.39	.98	2.47	.98	[illegible]
6. Instrument	84	2.52	.89	2.42	.85	[illegible]
7. Ranking	151	2.56	1.08	2.62	.94	.71

TABLE 4.6.—*Reliability of total scale scores for airplane commanders*

SCALE A AND B, SECOND AIR FORCE

Station	N	First rater, 1		Second rater, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Scale A:						
Alamogordo 3	23	12.61	2.57	12.65	2.69	0.44
Alamogordo 6	23	11.26	2.52	10.91	2.83	.80
Biggs	41	10.00	3.31	10.76	2.25	.61
Clovis	36	12.94	2.45	14.52	1.93	.34
Tucson	37	13.59	3.05	12.70	2.40	.50
Combined[1]	160	12.05	3.17	12.35	2.77	.55
Scale B:						
Great Bend	28	16.43	5.76	17.14	5.45	.82
Walker	49	17.53	6.43	16.50	5.57	.79
Albuquerque	42	15.55	5.74	17.46	5.42	.66
Pyote	46	18.89	5.07	19.81	5.98	.75
Combined[1]	165	17.22	5.88	17.78	5.78	.75

[1] By Fisher's z.

It is rather interesting that with the airplane commanders as well as with the other specialties and with the crew as a whole, the correlations between the ratings of the same individuals by different instructors are somewhat higher for over-all ratings than for ratings on specific traits or characteristics. This is shown by the fact that for Scales A and B the averages of the item reliabilities for the specific traits or characteristics (taking all crew positions and the crew as a whole) were 0.38 and 0.42, while the averages of the item reliabilities for the over-all and general ratings (all crew positions and crew as a whole) were 0.46 and 0.53 respectively.

The intercorrelations between ratings of the airplane commanders on the different traits and characteristics of the scale are shown in table 4.7. In

general ratings of interest in crew, of eagerness and of instrument flying seem to have less in common with the total set of ratings than do ratings of the other traits. As might be expected, the over-all or general ratings tend to have the most in common with the rest of the ratings. However, the differences are not large and all items have a great deal in common with the rest.

TABLE 4.7.—*Intercorrelations among item ratings on the airplane commander rating scale*

16th AND 17th WINGS—SECOND AIR FORCE

Airplane Commander—Scale A[1]

Item	1	2	3	4	5
1. Foresight	......	0.62	0.68	0.66	0.72
2. Interest in crew	......	......	.66	.58	.65
3. Formation flying	......	......	......	.62	.77
4. Instrument flying	......	......	......	......	.69
5. Lead possibility	......	......	......	......	......

[1] N = 160.

Airplane Commander—Scale B[2]

Item	1	2	3	4	5	6	7
1. Lead possibility	......	0.76	0.76	0.85	0.87	0.63	0.86
2. Formation flying	......	......	.64	.74	.78	.69	.78
3. Eagerness	......	......	......	.77	.77	.59	.76
4. Foresight	......	......	......	......	.86	.62	.85
5. Leadership	......	......	......	......	......	.66	.90
6. Instrument flying	......	......	......	......	......	......	.68
7. Ranking	......	......	......	......	......	......	......

[2] N = 170.

3. *Interview board ratings.*—At one station under the 16th BOT Wing in the Second Air Force aviation psychologists, training personnel and members of the Base Standardization Board cooperated in a special attempt to improve the estimation of the degree of proficiency of crew members completing operational training at that station. The Standardization Boards were interested in this problem because they had the mission of seeing to it that flight procedures taught in operational training conformed to all Army Air Forces directives. They also had the responsibility of checking on the performance of aircrew trainees to see that they were familiar with the standard flight procedures and flying regulations. After some discussion, it was decided to interview each aircrew officer completing operational training in a current class and on the basis of the interview and all other available data rate him on each of several traits considered important for success in combat. An interview board was therefore set up consisting of a member of the Base Standardization Board, an instructor of the specialty of the trainee being interviewed and an aviation psychologist. This board interviewed each airplane commander, copilot, navigator and bombardier completing operational training in a current class.

No set routine was established for the interview and the questions asked differed from one interview to another, although they were naturally di-

rected toward information related to the ratings to be made. After the interview, the boards rated each trainee on a 10-point numerical scale on each of the following traits: interest in flying, desire for combat, capacity to make decisions, getting along with others, educational background, and officer qualities. Decision as to the rating for a trainee on a particular trait was arrived at by consensus of board members. Where agreement was not reached, the average of separate ratings by the different board members was used. All available data from the Training Department were used in arriving at the rating as well as the impressions gained from the interview.

The intercorrelations between the ratings given by the Interview Board to 51 airplane commanders on the above 6 traits are shown in table 4.7. The surprising feature of these intercorrelations is the number of negative relationships shown, some of which are rather high. Although there was no definite evidence, it was suggested that the procedure of discussing each trait rating among the board members led to a greater degree of independence among the trait ratings, perhaps even to the extent that a high rating was given individuals in their most characteristic traits while low ratings were given in the remaining traits.

TABLE 4.8.—*Correlations between interview board ratings of six traits*

B-24 AIRPLANE COMMANDERS—SECOND AIR FORCE

Trait	Code[1]	A	B	C	D	E	F
Interest in flying	A	......	−0.24	0.25	−0.16	−0.10	−0.18
Desire for combat	B	......	......	−.51	.47	−.27	.41
Capacity to make decisions	C	......	......	......	−.53	.18	−.51
Getting along with others	D	......	......	......	......	−.23	.61
Educational background	E	......	......	......	......	......	.14
Officer qualities	F	......	......	......	......	......	......

[1] N = 51.

Since there seemed to be no simple way of estimating the degree of reliability, no such data were reported for the Interview Board ratings.

4. *Mutual ratings.*—As a part of the project in which aircrew officers completing operational training at one of the stations in the Second Air Force were rated by an Interview Board, each aircrew officer was asked to rate every other aircrew officer on his crew on the following traits: interest in flying; desire for combat; capacity to make decisions; getting along with others; educational background; and officer qualities. Each trait rating was made on a 10-point numerical scale, 1 being a poor rating and 10 the best rating. No data on reliability were reported.

Mission Ratings

Early in 1945 there was organized at Muroc AAF, California, a Lead Crew School under the joint supervision of the 21st Bomber Command and the Fourth Air Force. Crews thought outstanding in combat in the 21st Bomber Command were sent back to this school for special training for return to combat as lead crews. Believing that this school would afford an unusual opportunity to obtain information on lead crew proficiency, re-

search personnel of the Fourth Air Force spent considerable time observing training procedures, conferring with training personnel and studying available measures of crew proficiency. Two main types of measures of proficiency were found to be available. These were mission grades and ratings, and radar bombing scores. The latter are discussed briefly in a later section of this chapter, while a much more complete discussion is given in chapter 7, Bombardier.

After preliminary studies of mission grading and some discussions with training personnel, aviation psychologists in the Fourth Air Force were asked to collaborate with the training personnel at the Lead Crew School in the development of more adequate procedures for assessing the skill of a crew in accomplishing a mission. As a result of conferences and discussions scales were developed for rating each aircrew officer on each important activity he performed on a mission. In these scales, the instructor making the observations was supposed to record a rating of from 1, meaning "poor" to 5, meaning "excellent" for each of the mission activities on mission grading blanks provided for the purpose. There was a similar 5-point rating of over-all mission performance. In general, each trainee officer was rated by an instructor of his own specialty. In addition, each of the instructors in the different specialties made a 5-point over-all rating of the performance of the crew as a whole.

In practice, it was not possible to obtain ratings of all crew members on all missions. However, during the course of training at the Lead Crew School, it was usually possible to secure at least 6 mission ratings for each aircrew specialty. Usually a different instructor flew with a crew on each different mission, although occasionally an instructor might fly more than once with the same crew. Where an instructor repeated a rating on any crew, that rating was discarded in the statistical analyses of reliability. Thus all comparisons between odd and even mission ratings represented observations by different instructors. Reliability coefficients were obtained by comparing the averages of the ratings for odd and even missions. In general each mission blank yielded two scores. One was the total of the ratings on the separate activities of the mission and the other was the over-all rating of the individual on that mission. Table 4.9 presents the data on the reliability of the ratings of the airplane commanders. Data for the crew as a whole and for the other aircrew specialties are given in later chapters. Copies of the scales used in these mission ratings are given in appendix D.1.

TABLE 4.9.—*Odd-even mission reliability of mission ratings of airplane commanders*
LEAD CREW SCHOOL MUROC AAF—FOURTH AIR FORCE

Type of rating	June		July		August		Total	
	N	*r*	N	*r*	N	*r*	N	*r*
Average all items	23	0.39	27	0.45	21	0.53	71	0.45
Over-all rating	23	.37	27	.37	20	.71	70	.47

Proficiency Check List Ratings

1. *Crew proficiency check (VIIB)*.—Descriptive ratings, ratings of mission performance, and ratings on proficiency check lists are of course not necessarily mutually exclusive. Frequently parts of one device are quite similar to part or all of others of different types in terms of format, type of trait or characteristic rated, etc. In this sense, the Crew and Airplane Commander Proficiency Checks developed in the Third Air Force contained some items that were much like items on the Scale for Rating Mission Performance in the Fourth Air Force. However, the main emphasis in the two types of devices was rather different. In the Scale for Rating Mission Performance in the Fourth Air Force, the instructors were called upon to rate on a 5-point scale each of the most important activities of the crew member rated on that particular mission. In the Proficiency Checks, the instructor was called upon to check the presence or absence (or sometimes the amount of) certain items of behavior during the mission rated. This distinction is important and will be discussed further in the evaluation of the various criteria.

A great deal of time and effort went into the development of the Crew and Airplane Commander Proficiency Checks in the Third Air Force. In order to obtain firsthand information on the activities performed by the various crew members on the different missions, a research officer was attached to a CCTS class at one of the training stations of the Third Air Force. This officer flew each of the main types of required training missions at least once. As was described under Job Analysis in this Chapter, detailed records were maintained of all observable activities. Also studies were made of material obtained from conferences with instructors, analyses of training manuals and descriptions of training and combat situations where significant airplane commander performance was thought to have been demonstrated. From all of these sources were obtained items of behavior of airplane commanders on combat type missions which were thought to be both observable and important in indicating degree of proficiency. In constructing the check list, if possible, each item was set up in such a form that the observer merely made a check mark as to the presence or absence of the particular behavior on the mission being rated. However, certain important aspects of mission performance did not lend themselves to this type of report. Thus on some items the instructors were asked to estimate the frequency or quantity of some particular activity observed. With other items, the instructors were called on for a rating of the goodness or poorness of the behavior observed.

In this manner two check lists were developed by research personnel in the Third Air Force. The first, a preliminary form of an Airplane Commander Check List (B–24), was completed early in 1945. It was never used, however, since at that time the development of a check list for B–29 lead crew proficiency was given a higher priority.

The Crew Proficiency Check (VIIB) grew out of the earlier Airplane

Commander Check List (B-24). Items for the latter were revised to fit the B-29 training program and additional items were constructed covering the most important activities of crew members other than the airplane commander. Even in its final form, however, the Crew Proficiency Check (VHB) included a large proportion of items referring mainly to activities of the airplane commander, for 22 of the 49 check list items were of this type. This preponderance of airplane commander items was perhaps not undesirable in a device set up for the rating of crews as potential lead crew material, since the airplane commander was universally regarded as the most important person on the lead crew. In addition to the 49 check list items, there were included 2 general items, 1 calling for an over-all 5-point rating of the airplane commander and the other for a similar rating of the crew as a whole.

TABLE 4.10.—*Scores on crew proficiency check—VHB (B-29)*

THIRD AIR FORCE

Station	Weeks in training	N	Mean score	S. D.
MacDill	1–10	17	30.96	7.30
Gulfport	1–4	25	31.48	6.88
Gulfport	5–7	25	33.96	8.27
Chatham	1–4	19	33.26	7.94
Chatham	8–10	18	34.83	8.74
Barksdale	1–10	26	32.96	7.36
Total group	....	130	32.91	7.40

The final form of the Crew Proficiency Check (VHB) was administered to 175 B-29 crews in training at 4 Third Air Force Stations during the period from 5-15 August 1945. The check was administered to all crews with whom an instructor pilot rode during this period, regardless of mission or phase of training of the crew concerned. None of the crews was rated more than once so that it was not possible to obtain an adequate estimate of the reliability of these ratings. It had been planned that additional missions for each crew would be rated by an instructor other than the one making the original rating. Cessation of training after VJ-day made this impossible.

A scoring key was prepared for the Crew Proficiency Check (VHB) by asking experienced personnel to rate the different alternatives to each item in terms of degree of proficiency represented. With minor changes on the basis of material in training manuals and directives, these ratings were used to derive weights of 0, 1, 2, or 3 for each alternative under each item of the check list. Table 4.10 gives the means and standard deviations of the total scores on the check list for crews at each of the four stations where it was administered and where complete data were obtained.

Although a satisfactory estimate of reliability was impossible since the proficiency check was administered only once for each crew, an analysis was made of the consistency of the item scores for a single mission. A coefficient of internal consistency was computed by correlating the scores

obtained with 2 sets of 14 items selected by the research personnel as comparable in format and scoring. The uncorrected coefficient was 0.57. When this is corrected by the Spearman-Brown formula, the equivalent coefficient for the total scale of 49 items would be 0.82. It should be emphasized again that this is the consistency of a single observer's ratings of performance on a single mission.

2. *Instrument flying scale (B-17).*—Another device which attempted to obtain an objective evaluation of pilot proficiency was a scale for measuring instrument flying skill developed by aviation psychologists in the Third Air Force. This device consisted of a check list which covered the basic maneuvers of the white card instrument check, as outlined in AAF Regulation 50-3 (15 October 1943). No items were included, however, for the radio portions of the instrument check. Several types of items were included in the scale: quantitative items in which the observer checked such matters as "degree of bank," "time taken to attain a given change in altitude" during a specific maneuver, etc.; check list items in which the observer checked whether the pilot did or did not perform a certain act, such as "Trimmed plane—Yes——— No———"; and miscellaneous ratings involving judgments of the quality of certain specific activities of the pilot which did not lend themselves to quantitative description. A copy of the Instrument Flying Scale (B-17) is given in appendix E.1.

A preliminary form of the scale was administered by check pilots to 60 trainee pilots in the course of a routine white card instrument check. Item revisions were then made in the light of answers given by pilots with high, medium and low total scores on the scale. The completed scale was administered to 4 groups of pilots: 6 instrument check pilots; 19 flying instructors; 23 ground instructors, who were also pilots; and 20 combat crew trainee pilots. The differences between scores given pilots in the different groups were in the expected direction with the trainees receiving the lowest average score of 160 and the instrument check pilots receiving the highest average score of 177. The difference between the averages for these extreme groups was significant at the 1 percent level.

Each of the six instrument check pilots was scored on the instrument flying scale on five different occasions by each of two fellow check pilots. One of these observers remained the same for all five missions. The other observer was a different check pilot in each case. The average deviation of scores per check pilot was 5.6 when a different rater scored the five missions. When the same observer scored all missions, the average deviation was 5.3. The closing down of B-17 training prevented the use of the scale with larger groups of trainees. Thus no adequate data on reliability are available.

Self Ratings

Trait Ratings

Airplane commanders of one class of B-24 crews at a Second Air Force Station were asked to rate themselves on each of the following traits: interest in flying; desire for combat; capacity to make decisions; getting

ating with others; educational background; and officer qualities. The ratings were made on 10-point numerical scales with 1 representing a rating that was poor and 10 representing the best possible rating. No data on reliability were available.

In connection with descriptive ratings of lead crew proficiency obtained in the Second Air Force, there were included self-ratings of interest in lead crew assignment. These ratings were obtained in an effort to secure more information about the strength of motivation of the different crews and crew members. As with the descriptive ratings by instructors, the statements used with the self ratings differed considerably with Scales A and B, the two descriptive scales used.

The self ratings obtained from airplane commanders and other crew members in connection with administration of Scale A were in the form of five-point ratings under each of several different statements regarding degree of interest in lead crew assignment. In contrast to the rest of Scale A, the rating of one represented little interest and that of five, most interest. Although there were seven statements in all, data were analyzed only on the three following statements: "How interested are you in being a member of a lead crew?" "How interested are you in having your crew become a lead crew?" and "How interested is your crew in becoming a lead crew?" As would be expected there was a high degree of similarity in the answers of all crew members to the three different statements. For the last 2 of the 3 statements given above the correlations between ratings ranged from 0.67 to 0.80 for the different crew members of 156 B–29 crews. These coefficients probably represent an even higher degree of correspondence than indicated by their size, since the distributions of self-ratings were decidedly skewed. This is evidenced by the fact that the mean ratings were usually between four and five on a five-point scale, with standard deviations falling at about one.

With descriptive rating Scale B in the Second Air Force, a somewhat different self-rating of interest in lead crew assignment was obtained. Degree of interest was represented by ratings on a scale running from one, for little or no desire, to nine, for exceptionally strong desire for lead crew assignment. With Scale B the airplane commander rated only the strength of desire of his crew rather than that of himself. The other crew members rated their own desire for the assignment.

Airplane Commander Questionnaires

Another type of self-rating was that obtained from a paper and pencil inventory or questionnaire in which the individual described his own behavior. In connection with the development of procedures for the designation of certain airplane commanders as potential lead crew material, aviation psychologists in the Third Air Force developed such a questionnaire. In it, airplane commander trainees were asked to designate which of various types of behavior they customarily used in dealing with their crews. Much previous research and background material obtained from conferences with

experienced personnel was utilized in the construction of items for the questionnaire. A preliminary form was drawn up and administered to 89 airplane commanders at 1 station in the Third Air Force in June 1945. On the basis of the results from this administration the scale was revised and a final form containing 40 items was constructed. This form was administered to 175 airplane commander trainees at 4 heavy bombardment stations.

To obtain a scoring key, the research personnel consulted training personnel, training directives and manuals to obtain information on the basis of which they gave an arbitrary weight to each alternative for every item. The scores obtained with this key for the Airplane Commander Questionnaire are shown in table 4.11 for airplane commanders at each of 4 stations. The internal consistency or reliability of the scores on the questionnaire was determined by an analysis of the answers to the items, according to the Kuder-Richardson formula, $r_{tt} = \frac{n}{n-1} \cdot \frac{\sigma_{t^2} - n\bar{p}\bar{q}}{\sigma_{t^2}}$, the coefficients so obtained also being given in the table. It can be seen that the device had a moderate degree of internal consistency. A copy of the questionnaire is given in appendix C.1.

TABLE 4.11.—*Scores on airplane commander questionnaire*

B-29 CREWS—THIRD AIR FORCE

Station	Weeks in training	N	Mean score	S. D	Kuder-Richardson reliability
MacDill	10–12	45	39.6	6.7	0.64
Gulfport	8–10	41	38.9	6.0	.56
Chatham	8–10	46	43.6	6.6	.64
Barksdale	5–7	43	38.4	6.9	.68
All stations	....	175	40.2	6.7	.62

Rate of Training Accomplishment

One of the most important ways of estimating skill or proficiency of an individual or group is to measure the output per unit of time. Although this type of measure was not obviously and easily available in reference to the main objectives of operational training, reports of instructors and training personnel indicated a belief that the more skillful and proficient crews tended to accomplish their training requirements more quickly than less proficient crews. The outstanding crews were described as "efficient; they get things done." Such crews "breezed through their training." Comments such as these suggested that the rate at which a crew accomplished its missions might be a useful indication of its proficiency in operational training.

Early in 1944, more or less standard training procedures and procedures for reporting training progress were set up in all of the Continental Air Forces. It became possible then to estimate rather easily the rates at which different crews were completing their requirements in operational training. Thus research personnel of the Second Air Force were able to make detailed analyses of training accomplishment from the reports of training progress

on Form A 301, Crew Progress Chart—Flying Training. This chart contained separate sections for recording training accomplished in each of the main phases of training. (B-17 and B-24 training had three phases, while B-29 training was divided into two phases, with the charts differing accordingly). Another section was provided for recording the accomplishment of the AAF Training Standards requirements. Each section contained squares or spaces for recording hours flown as well as the specific requirements completed, each square representing 5 hours of flying by the crew concerned. When a crew completed a particular mission, the appropriate square or squares representing that mission and the particular requirements involved were blacked-in opposite the number designating that crew. Similarly the hours flown were entered by blacking-in the appropriate number of squares. By examining the charts for any particular date it was possible to determine very quickly, either for a single crew or for a class as a whole, the number of missions and requirements completed, number of hours flown, etc.

Great difficulty was encountered as soon as an attempt was made to estimate the reliability of mission accomplishment scores obtained from the crew progress charts. Operational training was, in a way, a series of unique events. Missions differed from each other and different phases of training represented quite different tasks. The different phases of training were carried on both concurrently and seriatim. In general, Phase I training was basic and dealt with checking out the individual crew members in their specialties. Hence this phase ten . to be completed early in training. However, long before a crew completed all of Phase I requirements it began to fly missions in Phase II, and in B-17 and B-24 training, missions of Phase III as well. Furthermore, the particular order in which missions were flown within each phase varied greatly from crew to crew. Actually the situation was such that although the order of accomplishment of missions varied greatly this order was not a chance matter and could not be treated statistically as if it were.

There were. of course, a great many uncontrolled variables that affected the amount of training accomplished by a crew during a given period of training. Some of these factors were specific to the crew. An example would be illness of one or more of the crew members. Others were characteristic of a particular class. Examples would be shortages of equipment and limitation of flying due to bad weather and the like. Still others were probably characteristic of the station. Such a factor would be a policy on the part of a training department to require all crews to fly certain missions during a particular week of training. Whatever the reasons, there were certainly wide differences in amount of training accomplished in a given period of time by crews in different CCTS classes.

The crew progress reports given the most detailed analysis in the Second Air Force were those for B-17 and B-24 crews showing the training accomplished at the end of the sixth week of training. The end of the sixth

week was chosen for this analysis since at that time most crews had not yet completed Phase I training, while at the same time a substantial number of requirements in Phase II and Phase III had usually been accomplished. The analysis included data for all crews in 24 CCTS (B-17 and B-24) classes at 10 different stations. The average number of training requirements (total for all phases) completed per class ranged from 76 to 108, with standard deviations of from 3.5 to 7.8. The total requirements at the time of this study numbered 127 as scored on the crew progress charts. Thus the mean percent of requirements completed at the end of the sixth week ranged from 60 to 84 for the classes studied. These percentages present a somewhat distorted picture of the condition of training at that time. Actually the requirements completed first were usually those most easily and quickly gotten out of the way. Those remaining tended to be requirements that would take the most time and be most difficult to accomplish.

The result of all this is that there was no simple statistical procedure for estimating the reliability of mission accomplishment scores. Three different ways of estimating this reliability were explored. First, there were compared the numbers of missions accomplished at the end of 6 weeks of training under each of the three phases of training. The following correlations were obtained between number of mission requirements completed in each of the three phases of training at the end of the sixth week: Phase I v. Phase II, 0.36; Phase I v. Phase III, 0.28; Phase II v. III, 0.34, giving an average intercorrelation between amount accomplished in the 3 phases of .33. Each coefficient was the average by Fisher's z, of the separate coefficients obtained with each of 24 CCTS classes at 10 different stations, and represented data from more than 1,200 crews. The chief objection to this estimate of reliability lies in the way training usually operated. If one crew concentrated on completing the activities in Phase II, it had less time to put in on activities in Phase I and Phase III. It was known that supervisors sometimes encouraged different crews to concentrate on different aspects of training so that the available equipment could be better utilized. To what extent this operated to lower the intercorrelations between scores for the three phases is not known.

A second procedure for estimating reliability was to compare the number of missions completed at different stages of training. This was done for four B-17 CCTS classes at two stations. Correlations between number of missions completed at the end of 4 weeks and number completed at the end of 6 and 8 weeks are given in table 4.12.

One important point should be made regarding the correlations between amount accomplished at the end of various amounts of training. It is that the amount accomplished at the end of the longer period of training represented the amount accomplished at the end of the shorter period plus whatever was accomplished in the time between. Hence the coefficients

listed in the table are spuriously high, due to the inclusion of the first score in the second.

To get around this difficulty the missions accomplished in successive two week periods were compared. Thus in table 4.12 are also shown the correlations between number of mission requirements completed from the fourth to the sixth week and the number completed from the sixth week to the eighth week. Taken at face value, the data in table 4.12 would indicate

TABLE 4.12.—*Correlations between number of missions completed at different stages of training*

B-17 CREWS—SECOND AIR FORCE

Number of missions completed fourth week, 1 v. number completed sixth week, 2

Station	Number of cases	Mean, *1*	S.D., *1*	Mean, *2*	S.D., *2*	r_{12}
Ardmore, 9–23	56	24.1	4.7	39.1	5.5	0.48
Ardmore, 11–1	56	25.3	3.3	40.8	3.7	.60
Ardmore, 12–10	56	18.0	3.6	30.4	6.3	.78
Alexandria, 10–6	54	24.5	2.8	35.6	3.6	.36

Number of missions completed fourth week, 1 v. number completed eighth week, 2

Station	Number of cases	Mean, *1*	S.D., *1*	Mean, *2*	S.D., *2*	r_{12}
Ardmore, 9–23	56	24.1	4.7	56.9	6.1	−0.21
Ardmore, 11–1	56	35.3	3.3	51.5	4.9	.46
Ardmore, 12–10	56	18.0	3.6	43.8	7.6	.53
Alexandria, 10–6	54	24.5	2.8	61.7	6.9	.24

Number of missions completed 4–6 weeks, 1 v. number of missions completed 6–8 weeks, 2

Station	Number of cases	Mean, *1*	S.D., *1*	Mean, *2*	S.D., *2*	r_{12}
Ardmore, 9–23	56	14.3	5.5	17.4	7.0	−0.06
Ardmore, 11–1	56	5.2	2.6	10.6	3.7	.08
Ardmore, 12–10	56	12.1	4.1	13.2	6.1	−.26
Alexandria, 10–6	54	10.5	3.9	11.7	5.8	−.47

that the accomplishment of missions score had a negative reliability. Of course, the main objection here lies in the fact that the crews that had completed most of their missions early in training usually had only the long and difficult missions remaining. Other crews that had completed less in their early weeks of training still had many easy missions which they could complete in a relatively short time later on in training.

The third type of analysis was to divide the training requirements into odd and even groups and compare the number of odd requirements completed with the number of even ones completed. The coefficients of reliability computed in this way are shown in table 4.13. The main objection to these figures is the same as that brought against the use of a split-half method in determining the reliability of any speed test. In a statistical sense the number of requirements completed during a particular period of training is like a score on a speed test.

Bombardment Scores

Although the main discussion of bombardment scores is found in chapter 7, Bombardier, some mention should be made here of their possible use as criteria of proficiency for airplane commanders. In operational training, as in bombardier training in the AAF Training Command, the average circular errors of bombs dropped on targets at bombing ranges were treated administratively as primarily the achievement of the bombardier. However, training personnel have usually emphasized that the skill of other

TABLE 4.13.—*Correlations between number of odd and even missions completed*

Data obtained at end of sixth week

B-17 CREWS—SECOND AIR FORCE

Station	Number of cases	Odd missions, 1		Even missions, 2		r_e
		Mean	S. D.	Mean	S. D.	
Ardmore 9-23	56	19.8	3.1	19.1	2.9	0.65
Ardmore 11-1	56	20.6	2.4	1[illegible].9	2.0	.44
Ardmore 12-10	56	15.5	3.6	14.7	3.1	.85
Alexandria 10-6	54	18.2	2.2	17.3	1.8	.66

crew members, particularly that of the airplane commander, have an important effect upon the accuracy of bombing. That this is true has been demonstrated rather conclusively in research on bombing accuracy at bombardier schools of the AAF Training Command, where it was found that the pilot was accountable for nearly as much variance in bombing scores as was the bombardier. With radar bombing scores the part played by crew members other than the bombardier was even greater, since a new crew member, the radar observer, carried on a vital part of the bombing operation. Thus, a brief discussion of average circular error and radar bombing scores is included in the discussions of criteria of proficiency of airplane commanders.

Average Circular Error

Because of the many known sources of error, the routine records of average circular error for combat crews in the Continental Air Forces were not well suited to serve as proficiency criteria. However, these scores did have some degree of reliability, since for 7 samples totaling nearly 600 B-17 and B-24 crews, the coefficients of reliability based on odd-even mission performance varied from 0.26 to 0.67. There was evidence that circular error averages based upon photographs of the bomb bursts were more reliable than those including both photographed and estimated circular errors. For further discussion of average circular error, see discussion under Criteria, chapter 7, Bombardier.

Radar Bombing Scores

Radar bombing of industrial targets in operational training simulated actual combat bombing conditions much more closely than did the actual release of bombs on targets at bombing ranges. The chief problem met in studies of radar bombing was that of securing adequate scoring of the bomb

run. A number of scoring methods were used during the period when studies of radar bombing scores were being made by aviation psychologists in the Continental Air Forces. Analysis of data obtained with the use of the different methods of scoring indicated that the best scoring procedure from the point of view of reliability was the radar scoring under Project 584, Radar Bomb Scoring. However, only a small amount of data were available, as this method of scoring was developed just prior to the close of hostilities. Odd-even mission reliability of this type of scoring radar bombing was 0.69. Further discussion of radar bombing scores is given in chapter 7, Bombardier, under Criteria.

Ground Training Performance

There was considerable difference of opinion among training personnel and officers with combat experience as to what extent performance in ground training was related to combat proficiency. The fact that it has been necessary to include this type of training even in the theaters of operations seemed to research personnel to justify some study of performance in ground school. Two types of records of performance were available. These were academic grades and scores on ground trainers.

Academic Grades

Academic grades were usually given each airplane commander in the various ground training courses required in operational training. The basis of grading, the type of grade and the significance attached to various grades varied with the instructor, course, class, station and air force. Sometimes the course grade was merely a rating by the instructor. Sometimes the course grade was the grade made on an achievement type examination given at the end of the course. Hence it is not possible to make many useful generalizations about the nature of these grades or about their usefulness as possible criteria of proficiency of airplane commanders.

An attempt was made to estimate the reliability of ground school grades at various stations in the Second Air Force. This was done by comparing the grades and scores achieved in successive periods of ground school training. Table 4..4 lists the reliability coefficients obtained in this way. These coefficients are probably spuriously high since it was not possible to deter-

TABLE 4.14.—*Coefficients of reliability of academic grades*

B-17 AND B-24 AIRPLANE COMMANDERS—SECOND AIR FORCE

	Number of cases	*r*
Engineering, first half v. second half of training	150	0.66
Bombardment, first half v. second half of training	110	.83
Communications, first half v. second half of training	55	.71
Gunnery, first half v. second half of training	56	.63
Average all courses, first half v. total training[1]	56	.76

[1] At this Second Air Force Station seperate grades were not recorded for the second half of training. Thus the grade for total training [illegible] the grades for the first half. Hence this coefficient has not been corrected by the Spearman-Brown formula. Even so it is probably spuriously high.

mine the extent to which scores in the first part of training influenced the scores given at later stages of training.

In addition to the above data, research personnel in the 16th Wing of the Second Air Force obtained designations of average ground school grades of crew personnel of each specialty in terms of the two categories of upper-half and lower-half of the class. No data on reliability of these grades were obtained.

Finally at March Field in the Fourth Air Force data were obtained on two tests given in ground school subjects: C-1 Autopilot Examination and Engineering Examination. No reliabilities were determined for these scores.

Ground Trainer Scores

1. The *Link Trainer* was the most important of the ground trainers worked with by airplane commanders, although in operational training they were also given training on other types of trainers. The Link Trainer was a device consisting of a cockpit with instrument panel and controls roughly similar to those found on a relatively simple type of training plane. The device was so constructed that manipulation of the controls resulted in changes in the attitude and position of the cockpit and changes in readings of instruments which corresponded closely to changes which would result from similar operation of controls in real flight. The trainer also made a record of the simulated flight path of the pilot including all of the important variables of flight.

There was great variation between stations in the amount of type of training given and in the records maintained thereof. Usually for each training session or "mission" a score was recorded which represented the judgment of the instructor as to the skill shown by the pilot. This score was most often a three-point rating and was based on a judgment of the degree to which the pilot maintained the briefed course according to the recorded flight path. The reliability of Link Trainer scores is indicated by the data presented in table 4.15, where the average scores obtained on odd and even missions are compared. For the data in the table the number of missions per pilot varied from 6 to 10. In the case of the B–17 data, all cases were obtained from the same station and a single coefficient of reliability was computed. With the B–29 data, separate reliability coefficients were computed for each class and station. The different coefficients were then averaged by Fisher's z to yield the coefficients given in the table.

Similar Link Trainer ratings were available at two stations in the Fourth

TABLE 4.15.—*Reliability of link trainer scores*

SECOND AIR FORCE

Type of plane	N	Odd scores		Even scores		r_{tt}
		Mean	S. D.	Mean	S. D.	
B–29	676	1.93	0.26	1.95	0.31	0.49
B–17	391	1.98	0.19	1.99	0.19	.13

Air Force. The odd-even mission reliability, based on ratings of 5 missions, was found to be 0.63 for 56 airplane commanders. In addition to the ratings a Link Trainer Examination of 30 objective-type items was administered at the outset of training at 1 of these stations. Data on this test were compared with average ratings on the Link Trainer missions, the latter being divided into two categories of upper and lower half of the group. The biserial coefficient of correlation between examination score and Link Trainer rating was 0.46 for 52 airplane commanders.

A rather different type of measure of proficiency was also sometimes available in connection with Link Trainer performance. This was the number of Link Trainer hours taken by a pilot to complete the minimum requirements. Of course, this measure was meaningless at those stations at which there were no minimum proficiency standards and at stations where the requirements were solely in terms of hours of practice. As with other similar measures (e.g. Rate of Training Accomplishment, etc.) it was not possible to obtain any estimate of the reliability of this score.

2. *Miscellaneous trainers.*—Airplane Commanders were also required in operational training to practice on various navigation and bombing trainers. The amount and nature of such training varied even more than did training on the Link Trainer. Frequently the training represented only a familiarization with the equipment involved. Often no attempt was made to assess the performance of the airplane commanders. However, sufficient trainer practice was required at one Second Air Force station to provide data for computation of reliability of certain trainer scores. The coefficients of reliability for Dead Reckoning and Celestial Navigation Trainer Scores of airplane commanders were found to be 0.37 and 0.18 respectively, based on the correlation between average scores for odd and even missions. With the Celestial Navigation Trainer only two missions were available while on the Dead Reckoning Trainer there were at least four missions. At AAF Muroc (prior to establishment of the Lead Crew School) Fourth Air Force, circular error scores were obtained for a number of airplane commanders on the A-2 Synthetic Bomb Trainer. No reliabilities were computed for these scores.

Flying Evaluation Board Reports

The principal function of Flying Evaluation Boards was to investigate and appraise aircrew officers who had refused to fly, who were charged with some type of incompetence, or who were believed not to meet the medical standards required for flight status. FEB reports of varying degrees of completeness were usually available in each of the Continental Air Forces. Sometimes these reports included much pertinent biographical material, testimony of associates and supervisors and other material as well as the action of the Board. In other cases the reports included only the bare statement of Board action with a reason for the action stated in a single word or short phrase. The major difficulty in the use of reports of

Flying Evaluation Boards as criteria of proficiency was the lack of standardization in methods of investigation and evaluation, and the brevity of some of the reports. Usually, however, the reports at least permitted classification of board action into three main categories of reasons for action: personality defect (principally fear of flying); lack of proficiency; and physical disability. The percentages of cases falling into the various categories are given in an earlier section of this chapter under Job Specifications.

Since final board action was in each case a unique event, it was not possible to make any estimate of the reliability of this criterion.

Aircraft Accidents

Pilots having aircraft accidents directly attributable to lack of skill in handling the aircraft may obviously be considered as showing a relatively low degree of proficiency—at least at the time of the accident. Hence research personnel in each of the Continental Air Forces made some study of aircraft accidents and attempted to use their occurrence as a criterion of lack of proficiency in validation studies. The relative frequencies of various types and causes of accidents in heavy bombardment training are shown in tables 4.16 through 4.19. The data from the Second Air Force were based on the records of the 296 bombardment aircraft accidents occurring in the command between 1 October 1944 and 1 April 1945. The Fourth Air Force data included 165 B–24 accidents reported during January and February 1945.

The data in the tables are more or less self-evident. It might be pointed out that while there were marked differences between the Second and Fourth Air Force data in the percents of different types of accidents(table

TABLE 4.16.—*Rate of bombardment aircraft accidents per 1000 hours flown*

OCTOBER 1944 TO APRIL 1945—SECOND AIR FORCE

Type of aircraft	Total rate	Fatal rate
B-17	26	6
B-24	28	7
B-29	50	17

TABLE 4.17.—*Percents of bombardment aircraft accidents of different types*

SECOND AND FOURTH AIR FORCES

Nature of accident	Percent of all bombardment accidents, Second Air Force	Percent of all bombardment accidents, Fourth Air Force
Emergency landing	17	[illegible]
Normal landing	17	[illegible]
Take-off	7	[illegible]
Mid-air collision	2	10
Collision with terrain	6	8
Emergency bail-out	....	7
Fire in air	10	....
Structural failure	1	....
Taxiing	21	6
Miscellaneous	19	0
Total	100	100

4.17), the percent attributable to pilot error (table 4.18) was the same in both cases. The differences in nature of the accidents (table 4.17) may well have been due to differences in the type of plane involved. The majority of the Second Air Force accidents were B–29 accidents while those in the Fourth Air Force were B–24 accidents.

It was only rarely that an airplane commander had more than one accident during operational training. As rough evidence here might be cited the

TABLE 4.18.—*Percents of bombardment aircraft accidents of different causes*

SECOND AND FOURTH AIR FORCES

Cause of accident	Percent of all bombardment accidents, Second Air Force	Percent of B-24 accidents, Fourth Air Force
Personal error:		
Pilots	33	
Maintenance personnel	1	
Supervisory personnel	2	
Other personnel	15	
Total	51	47
Materiel failure:		
Power plant	15	
Landing gear	9	
Other parts	3	
Total	27	46
Undetermined and miscellaneous	22	7
Grand total	100	100

TABLE 4.19.—*Percents of different errors in flight technique causing accidents*

88 B-24 PILOTS—FOURTH AIR FORCE

Type of error in flying technique	Percent of pilot error accidents
Operation of flight controls	29
Power-plant procedures	20
Supervision of crew	12
Operation of auxiliary controls	11
Judgment	9
Radio technique and copilot procedure	8
Observation and orientation	7
Safety violations	4
Total	100

fact that out of the 161 B–24 pilots in the Fourth Air Force who had accidents in January and February 1945, only 4 had 2 such accidents. Suffering an aircraft accident was almost a unique (in a statistical sense) event for an airplane commander. Thus it was not possible to determine the reliability of the occurrence of aircraft accidents as a criterion of proficiency.

Assignment as Airplane Commander or Copilot

Whether a particular graduate from Advanced Pilot School was assigned to further training as an airplane commander or as a copilot was not, strictly speaking a criterion of proficiency in operational training. How-

ever, the fact that each bombardment crew studied had an airplane commander and a copilot made operational training a convenient source of data for a study of the validity of selection procedures in predicting such assignment. There were naturally no data on the reliability of this assignment. Available information on the procedures sometimes used would lead one to the conclusion that the matter was frequently decided largely on the basis of administrative convenience.

Intercorrelations Among Criteria

The exploratory nature of many of the studies of criteria made by aviation psychologists in the Continental Air Forces, and the wide differences between stations in training procedures and records maintained, made systematic study of the relationships between all criteria impossible. In the paragraphs which follow are presented what data were available. Most of them are fragmentary and the numbers of cases available for the different comparisons are rather small.

Correlation Between Ratings

The intercorrelations between various types of ratings made on B–17 airplane commanders in the Second Air Force are given in table 4.20. The Efficiency Rating, Form 66-2, Officer Quality Rating and Technical Aircrew Skill ratings were five-point ratings ranging from "poor" to "superior." The AFTRC Form 2 ratings were three-point ratings with the middle rating being "average." Rank in Class was a rating in which instructors listed the airplane commanders in a class in the order of their expected proficiency in combat. The coefficients of correlation are given above and to the right of the diagonal in the table, while the numbers of cases included in each coefficient are designated below and to the left. It should be stated in this connection that the same personnel usually did most of the ratings for a given set of crews.

TABLE 4.20.—*Correlations between different ratings of airplane commanders*[1]

B–17 AIRPLANE COMMANDERS—SECOND AIR FORCE

Type of rating	Code	A	B	C	D	E	F	G
Efficiency rating, form 66–2	A	....	....	....	0.36	0.45	0.50	0.6[illegible]
Officer quality rating	B	....	....	0.44	.64	.48	.49	.....
Technical aircrew skill rating	C	....	55	....	.58	.52	.67	.....
AFTRC Form 2:								
Military discipline	D	82	21	21	....	.50	.25	.23
Manner of performing duties	E	82	21	21	103	....	.45	.40
General aircrew ability	F	82	21	21	136	136	....	.38
Rank in class	G	167	....	....	82	82	82	.....

[1] Entries below the diagonal are the numbers of cases for the corresponding correlation coefficients.

A number of different types of ratings of B–24 airplane commanders were also made in the Second Air Force where each rating was made independently by a different observer or group of observers. The intercorrelations between such ratings are given in table 4.21. The interview rating was an average of 10-point ratings for 6 traits made by a board of 2 training of-

ficers and an aviation psychologist who interviewed each of the airplane commanders. The ground school rating was the average of the ratings of the airplane commanders for all academic courses. The flight surgeon's rating was a 3-point rating of which the middle rating was "average." The crew and self-ratings were averages of 10-point ratings on 6 traits by fellow crew members and by the airplane commander himself. Rank in class was the listing of the airplane commanders in the order of their expected effectiveness in combat. In the table, the coefficients of correlation are given to the right and above the diagonal and the numbers of cases included in each are designated below and to the left.

TABLE 4.21.—*Correlations between different ratings of airplane commanders*[1]

B-24 AIRPLANE COMMANDERS—SECOND AIR FORCE

Type of rating	Code	A	B	C	D	E	F
Interview rating	A	......	0.06	0.21	0.17	[illegible]	0.[illegible]
Ground school rating	B	[illegible]	......	−.16	−.[illegible]	.14	.[illegible]
Flight surgeon's rating	C	52	52	......	−.13	.[illegible]	.13
Crew rating	D	52	52	52	......	−.[illegible]	−.[illegible]
Self rating	E	52	52	52	52	......	−.[illegible]
Rank in class	F	[illegible]	[illegible]	52	52	52	......

[1] Entries below the diagonal are the numbers of cases for the corresponding correlation coefficients.

The correlations between self-ratings of interest in lead crew assignment and over-all ratings of lead crew proficiency for B-29 airplane commanders rated with Scales A and B in the Second Air Force are shown in table 4.22. For Scale A, the three statements required ratings by the airplane commander of his desire for his own and for his crew's assignment in lead crew position and rating of his crew's desire for such assignment. With Scale B only the latter rating was obtained. The coefficients in the table show that, especially for Scale B, there was a slight tendency for airplane commanders who reported high degree of interest in lead crew assignment to be regarded by instructors as good prospects for such assignment.

TABLE 4.22.—*Correlations between interest and over-all rating of lead crew proficiency*

B-29 AIRPLANE COMMANDERS—SECOND AIR FORCE

Type of statement of interest	N	Degree of interest, 1		Rating, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Scale A:						
Desire for own assignment	156	4.74	0.53	2.41	0.60	0.12
Desire for crew's assignment	156	4.76	.49	2.41	.60	.14
Desire of crew for assignment	156	4.51	.78	2.41	.60	.63
Scale B:						
Desire for crew assignment	156	6.97	2.28	2.46	.74	.29

The Airplane Commander Questionnaire in the Third Air Force represented a kind of self-rating. Both it and the Crew Proficiency Check (VHB) were administered for 175 B-29 crews at 4 Third Air Force stations. There were 12 items of the Crew Proficiency Check for which there were directly parallel or analagous questions in the Questionnaire. It was, therefore, of some interest to compare the scores made by the airplane commanders on the 12 items of the Questionnaire with the ratings given them

on the corresponding items of the Crew Proficiency Check. The correlation between scores on the two sets of items was 0.36 for 175 airplane commanders.

Correlations Between Ratings and Bombing Scores

Data on the crew average circular error for practice bombing of targets on bombing ranges were available for the crews of 434 pilots of 8 CCTS (H) who were ranked in their flights by their instructors in terms of their expected effectiveness in combat. The correlation between rank in flight and crew average error was determined separately for each class. These coefficients ranged from —0.17 to 0.37 and when averaged by Fisher's *z* gave a coefficient of 0.05 for the total group. All coefficients were first adjusted in sign so that a positive coefficient represented correlation with goodness of performance.

The correlations between average circular error and ratings of B–29 airplane commanders on descriptive ratings scales A and B in the Second Air Force are given in table 4.23. Although instructors reported that the airplane commander played an important part in the bombing operation, the data in table 4.23 show that the bombing accuracy achieved was given little weight in their ratings of airplane commanders.

Similar correlations with radar bombing circular error T-scores were available for airplane commanders rated with Scale B in the Second Air Force. The radar bombing T-scores represent the combined scores from three types of radar bombing scoring procedures, each type being converted

TABLE 4.23.—*Correlations between ratings of airplane commander and crew bombing scores*

B–29 CREWS—SECOND AIR FORCE

Station	N	Average circular error, *1*		Average rating per item, *2*		r_{12}
		Mean	S. D.	Mean	S. D.	
Rating scale A:						
Alamogordo 3	23	340	63	[illegible].52	0.46	0.16
Alamogordo 6	23	331	72	2.72	.51	—.22
Biggs Field	41	266	60	2.[illegible]	.52	—.01
Clovis	36	242	47	2.73	.35	.46
Tucson	37	282	59	2.61	.52	.21
Combined	160	284	69	2.43	.55	[1].15
Rating scale B:						
Albuquerque	45	265	54	2.28	.6[illegible]	—.18
Great Bend	44	210	53	2.42	[illegible]	—.01
Pyote	49	242	54	2.69	.7[illegible]	.23
Combined	138	239	57	2.47	0.76	[illegible]

Station	N	Radar bombing T-scores, *1*		Average rating per item, *2*		r_{12}
		Mean	S. D.	Mean	S. D.	
Rating scale B:						
Albuquerque	21	50.0	4.40	1.98	0.49	0.04
Great Bend	31	50.1	5.32	2.44	.90	—.20
Walker	34	49.8	4.14	2.49	.77	—.22
Combined	86	49.99	4.66	2.35	.79	[1]—.15

[1] Combined by means of Fisher's *z* technique.

separately to corresponding T-scores. These data are also presented in table 4.23. Here again there is no evidence that accuracy in bombing was given any weight in the ratings of the airplane commanders.

Correlations Between Ratings and Miscellaneous Criteria

Ground school grades were available for B–29 airplane commanders rated with descriptive rating Scale A in the Second Air Force. These grades were given in terms of upper-half or lower-half designations for all airplane commanders rated with Scale A. The biserial coefficient of correlation between ratings and ground school grades was 0.06 for 160 airplane commanders. Plainly performance in ground school did not greatly influence the ratings given the airplane commanders by flying instructors.

The correlations between ratings of B–29 airplane commanders on Scale A in the Second Air Force and various crew accomplishment of training scores is given in table 4.24. All coefficients shown were based on an average by Fisher's z of separate coefficients for each class and station. Apparently efficient accomplishment of training requirements did influence somewhat the ratings of airplane commanders.

TABLE 4.24.—*Correlations between descriptive ratings and crew accomplishment of training*

SCALE A—154 B–29 AIRPLANE COMMANDERS—SECOND AIR FORCE

Type of accomplishment score	Training accomplished, I		Ratings, I		r_{12}
	Mean	S. D.	Mean	S. D.	
Missions completed in 6 weeks	36.8	5.62	2.41	5.36	0.26
Hours of flying in 6 weeks	11.5	2.82	2.41	5.36	.20
AAF minimum standards, 6 weeks	38.7	17.1	2.41	5.36	.26

In table 4.25 are given the intercorrelations between various ratings and test and trainer scores for B–24 airplane commanders in the Fourth Air Force. The over-all ratings were on the usual five-point military scale, and the Link Trainer Ratings were on a three-point scale. Examination Scores were on the usual academic scale of 100 points. Units used in A-2 Circular Error scores are not known. The highest intercorrelations were found between C-1 Autopilot and Engineering Examinations, between Link Trainer Rating and Link Trainer Examination, and between Over-all Rating and Engineering Examination.

The correlation between ratings on Descriptive Rating Scale B and hours taken to complete Link Trainer requirements was 0.29 for 41 airplane com-

TABLE 4.25.—*Intercorrelations among criteria of proficiency*

B–24 AIRPLANE COMMANDERS—FOURTH AIR FORCE

Type of criterion	Code	Intercorrelations								S. D.
		A	B	C	D	E	F	N	Mean	
Overall rating	A		0.14	−0.01	0.15	0.44	−.10	193	5.29	4.53
Link trainer rating	B			.46	.10	.12	−.08	193	11.9	1.69
Link trainer examination	C							52	79.5	9.93
C-1 autopilot examination	D					.66	−.14	40	87.3	6.77
Engineering examination	E						.19	40	81.0	9.30
A-2 circular error	F							40	12.7	1.97

manders at one Second Air Force Station. At another station approximately half of the airplane commanders failed to meet the special Link Trainer requirements set up there. The biserial coefficient of correlation between ratings on Scale B and completion of Link Trainer requirements was 0.26 for the 46 airplane commanders involved in the study.

Evaluation of Criteria

In the evaluation of the various types of proficiency criteria, it is useful to consider each of them in terms of certain features desirable in a criterion. In order to be as brief as possible, the total group of criteria of airplane commander proficiency described in the preceding sections will be discussed below under each of five main desirable features.

Nature and Distribution of Scores

Many of the criteria of airplane commander proficiency meet this requirement in a satisfactory manner. However, certain of them do not. For example, the routine efficiency ratings, the upper and lower half ratings of proficiency in the Fourth Air Force and all of the three-point ratings including Link Trainer ratings provide too little discrimination between individuals to be of maximum value. For a somewhat different reason the elimination of airplane commanders by Flying Evaluation Board Action and the occurrence of Aircraft Accidents Due to Pilot Error are also unsatisfactory. Here the chief difficulty is the small percentage of airplane commanders that is involved. Finally, the self-ratings of Interest in Lead Crew Assignment of Scale A in the Second Air Force produced distributions of scores that were too skewed for practical use as criteria.

Reliability or Consistency

Under this requirement, the chief problem has been to obtain evidence upon which to base a judgment. For none of the following types of ratings was there any data on reliability obtained: 3-point Over-all Ratings in the Second Air Force; Interview Board, Mutual, and Self-Ratings of B-24 airplane commanders in the Second Air Force; and all Check List Ratings in the Third Air Force. However, if the criteria had been considered satisfactory in other ways, most of these reliabilities could have been determined fairly easily. In the case of the Third Air Force Check List Ratings the closing down of crew training prevented the securing of the desired data.

In contrast, Rate of Training Accomplishment, Hours to Complete Link Trainer Requirements, Flying Evaluation Board Action and Aircraft Accidents were not susceptible to analysis in terms of reliability under conditions of operational training. Of moderate and fair degree of reliability were the Descriptive Ratings of Lead Crew Proficiency in the Second Air Force, the Mission Ratings in the Fourth Air Force and Radar Bombing Scores. Average Circular Error for Practice Bombs, Academic Grades and certain Ground Trainer Scores had some reliability but not as much as would be desirable in an acceptable criterion score.

Discrimination Between Airplane Commanders of Known Proficiency

No evidence under this heading was obtained for criteria of airplane commander proficiency in the Continental Air Forces. Because of their nature, even if ratings did discriminate between individuals known to differ in proficiency, this fact would be no point in their favor. The degree of proficiency of the groups concerned would have to be unknown to those making the ratings—a situation that would be practically impossible in operational training.

Breadth of Activities Covered

Superficially at least the Over-all Judgment Ratings and the Flying Evaluation Board Actions would seem outstanding in terms of this feature of criteria. However, a little further thought would lead to some question here. The extent to which the individual rater really canvassed all aspects of performance of the airplane commanders being rated is open to doubt. Decision by a board of several individuals might be expected to be somewhat better in this respect. In any case how thorough a consideration was given all aspects of the job in any set of ratings was not known. All ratings involving ratings of several specified traits including the Check List type of ratings probably covered fairly well those aspects of airplane commander proficiency actually specified in the instrument concerned. The extent to which they represent complete coverage depends upon how well the particular set of traits or characteristics covers the job of airplane commander. In the case of criteria reported in this chapter, the sets of traits and characteristics covered by the various instruments and rating devices were those provided by groups of experts and may be presumed to represent fair coverage of the important aspects of the job.

The accomplishment of training criteria are open to the objection that speed in getting things done is not the sole aspect of performance that is important. Ground Training Criteria, Aircraft Accidents and Bombing Scores represent only part of the important aspects of the job of airplane commander.

Objectivity

Objectivity is the main feature of proficiency criteria wherein ratings of most types fail miserably in meeting the need in research on selection and training. In this respect, over-all numerical ratings are probably the worst. With descriptive ratings the attention of the rater is at least directed toward different aspects of the proficiency of individuals being rated. It is interesting that where reliability is thought of as agreement between ratings of different observers, such agreement may also be obtained by similiarity of bias or prejudice. Logically best among the various types of ratings are those that call for evaluation of observed behavior of the person being rated. Mission Ratings are thus probably superior to general Descriptive Ratings. And best of all in this respect are the Check List Type ratings

such as those developed in the Third Air Force. It is probable that Academic Grades (except those based largely upon examination scores) and Flying Evaluation Board Action are also open to a considerable degree of subjectivity. The remaining criteria studied would seem to have satisfactory objectivity.

Susceptibility to Miscellaneous Sources of Error

There could be almost endless discussion of all of the possible sources of error for the various criteria of airplane commander proficiency studied. Probably the most important point to be made here is one that has been repeated throughout the discussions of the various criteria. This is that training procedures and records kept varied with the instructor, flight, class, station, and air force. Attempts to specify rigidly the conditions of training met with failure for all sorts of reasons. Differences in experience on the part of instructors, changes in equipment, weather, personnel shortages, changes in commitments, differences in experience on the part of trainees, changing requirements in combat, etc., are but a few of the sources of variation. Where a criterion involved routine records, such factors as carelessness, desire to make a good showing and the like were important sources of error. Sheer pressure of time and overworking of training personnel played their part. Unfortunately the more objective types of criteria were probably as much affected by these miscellaneous sources of error as were ratings and other subjective criteria.

Recapitulation

In summary, it is believed that Mission Ratings and Check List Ratings are the most likely of the rating procedures to be useful in validation and other research. Unfortunately data on the reliability of the Check List Ratings were not obtained and only an hypothetical judgment can be made in this case. Ground Trainer Scores, Rate of Training Accomplishment, Aircraft Accidents, and Flying Evaluation Board Action would seem the best of the remaining criteria, all things considered. Bombing scores might be included except that they are probably determined to an equal or greater extent by the proficiency of other crew members. Finally, because of the low degree of intercorrelation between the ratings on different traits, the Interview Board Ratings would be interesting to study further.

VALIDATION

Stanine Validity

Ratings of Over-all Proficiency

In table 4.26 are shown the correlations between the augmented pilot stanine of B-17 and B-24 airplane commanders in the Second Air Force and various types of ratings of over-all proficiency. The efficiency rating, Form 66-2, rating of attitude toward work and ratings of technical aircrew

skill and officer quality were all five-point numerical ratings. The general ability rating on AFTRC Form 2 and the flight surgeon's ratings were three point ratings. Rankings by instructors were listings of airplane commanders in their flights in the order of their expected effectiveness in combat. None of the coefficients differs significantly from zero, although all but two are positive. Considerable additional data were obtained on the validity of the pilot stanine in predicting rankings by instructors of airplane commanders in their flights and classes. For 13 CCTS classes with a total of 670 airplane commanders (including data shown in table 4.26) the correlations between rankings and augmented pilot stanine when averaged by Fisher's z gave a coefficient of 0.03.

TABLE 4.26.—*Correlations between aptitude scores and over-all ratings*

B-17 AND B-24 AIRPLANE COMMANDERS—SECOND AIR FORCE

Criterion	Number of cases	Stanine, 1		Rating, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Efficiency rating, Form 65-2	134	6.31	1.89	46.9	12.8	−0.06
Attitude toward work	189	6.54	1.60	2.07	.31	−.04
Rankings by instructors	134	6.31	1.89	49.2	19.6	.08
Technical aircrew skill	33	6.06	1.83	54.4	8.8	.08
Officer quality	33	6.00	1.83	51.6	15.5	.18
AFTRC Form 2, general ability	51	6.47	1.72	54.4	9.8	.13
Flight surgeon's rating	54	5.96	1.43	2.67	.58	.16

Descriptive Ratings

The correlations between the stanines of B-24 airplane commanders and upper-half and lower-half ratings of airplane commander proficiency by instructors in the Fourth Air Force are shown in table 4.27. Since this was a dichotomous criterion, all coefficients are in biserial form. All three stanines showed a statistically significant correlation with the ratings, that for the bombardier stanine being significant at the one percent level.

TABLE 4.27.—*Validities of stanines of airplane commanders in predicting ratings by instructors*

B-24 TRAINING—FOURTH AIR FORCE

Type of stanine	N_t	N_l	M_t	M_l	SD_t	r_{bis}
Bombardier stanine	130	59	6.18	5.68	1.71	**0.33
Navigator stanine	130	59	6.06	5.64	1.76	*.27
Pilot stanine	129	58	6.20	5.83	1.53	*.27

t = Total group of airplane commanders.
l = Airplane commanders rated in lower half.

The correlations between augmented pilot stanines of B-24 airplane commanders and Interview Board and Mutual ratings on 10-point scales on 6 traits are shown in table 4.28. The correlation between stanines and Interview Board Ratings was significant at the five percent level. In the case of the Mutual Ratings the relationship was negative and not significant.

The degree to which the various stanines predicted descriptive ratings of lead crew proficiency of B-29 airplane commanders in the Second Air

Force is shown in table 4.29. The table includes data from both Scale A and Scale B. The rating score used was the average rating per item for the combined ratings of two raters on all items of each scale. These data were based on two sets of ratings because it was believed that some of the errors

TABLE 4.28.—*Correlations between augmented pilot stanines and sum of trait ratings*

B-24 AIRPLANE COMMANDERS—SECOND AIR FORCE

Type of rating	N	Pilot+Cr. stanine, 1		Ratings, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Interview board	54	5.96	1.43	5.91	3.11	*0.25
Mutual rating	54	5.96	1.43	5.18	1.69	−.07

of subjective bias might be minimized by this means. It will be noted that while Scale A was administered for 160 airplane commanders and Scale B for nearly 200 airplane commanders, only 73 and 86, respectively, were found to have been given the Aircrew Classification Tests. Thus the airplane commanders for whom data are shown here may represent a selected group, the selection of which may have involved some type of bias. One of the most interesting features of the results is the tendency for the bombardier stanine to predict the ratings better than the pilot stanine. This is in agreement with the data from the Fourth Air Force.

TABLE 4.29.—*Correlations between stanines of airplane commanders and descriptive ratings*

SCALES A AND B—B-29 CREWS—SECOND AIR FORCE

Station	N	Rating		Bombardier stanine			Navigator stanine			Pilot Cr. stanine		
		Mean	S. D.	Mean	S. D.	r	Mean	S. D.	r	Mean	S. D.	r
Scale A:												
Alamogordo 3	13	2.60	0.36	6.38	1.44	0.21	6.23	1.80	[illegible]	6.69	1.73	0.15
Alamogordo 6	9	2.41	.23	6.00	1.25	.07	5.34	1.70	[illegible]	7.44	1.64	.25
Biggs Field	15	2.10	.31	6.20	1.94	−.04	6.06	1.95	.02	[illegible]	1.30	.13
Clovis	20	2.78	.35	5.70	1.90	**.68	5.50	1.86	.25	[illegible]	1.81	.32
Davis-Monthan	16	2.73	.24	6.56	1.62	.23	5.75	2.08	.33	[illegible]	1.50	.17
All stations	73	2.55	.36	6.11	1.77	[1]**.33	5.97	1.89	[1].17	6.92	[illegible]	[1].19
Scale B:												
Albuquerque	27	2.34	.55	5.89	1.85	−.06	5.48	1.29	−.24	6.07	[illegible]	[illegible]
Great Bend	12	2.61	.66	6.08	1.98	−.31	6.17	1.82	−.23	6.83	[illegible]	−.20
Pyote	35	2.90	.63	5.91	1.56	*.43	5.63	1.55	.10	6.40	1.44	.20
Walker	12	3.06	.74	6.17	1.08	.08	5.50	1.67	−.14	6.58	1.67	.31
All stations	86	2.70	.69	5.97	1.67	[1].14	5.64	1.55	[1]−.09	6.38	1.52	[1].06

[1] Averaged by means of Fisher's *z* technique.

In connection with the above data the question arises whether the relative validity of the pilot stanine in prediction of the ratings of a single rater is really less than its validity in prediction of the combined ratings of two observers. The data on this point were inconclusive.

Mission Ratings

It was not possible to determine the validities of the various stanines of B-29 airplane commanders in predicting average mission grades at Lead

Crew School at AAF Muroc in the Fourth Air Force since few of them had been given the Aircrew Classification Tests in the AAF Training Command.

Check List Ratings

No data were obtained on the degree to which stanines of airplane commanders were related to scores on the Crew Proficiency check (VHB) and on the Scale for Instrument Flying developed in the Third Air Force.

Self-Ratings

The correlation between the pilot stanine of 54 B-24 airplane commanders and their self-ratings on a 10-point scale on a total of 6 traits was found to be —0.16. This negative coefficient was not statistically significant.

The Airplane Commander Questionnaire developed in the Third Air Force was administered to 175 B-29 airplane commanders at 4 stations. For the 78 airplane commanders for whom data were available the correlations between pilot stanine and score on the Questionnaire are shown in table 4.30. The interpretation to be placed upon the negative relationship found is not clear. A possible explanation would be that the poorer airplane commanders gave themselves the benefit of any doubt and claimed acceptable alternatives in the test more often than did the better airplane commanders. At the moment sufficient data are not available to determine what factors were involved.

TABLE 4.30.—*Correlations between pilot stanine and score on airplane commander questionnaire*

B-29 TRAINING—THIRD AIR FORCE

Station	N	Questionnaire score, *1*		Pilot stanine, *2*		r_{12}
		Mean	S. D.	Mean	S. D.	
Barksdale	15	60.3	12.50	6.6	1.54	−0.23
Gulfport	25	66.2	10.00	6.1	1.83	−.09
Chatham	18	74.1	6.36	6.3	1.99	−*.56
MacDill	20	68.5	8.32	7.1	1.87	.03
All stations (average by Fisher's z)	78	64.91	10.44	6.5	1.87	−*.21

Accomplishment of Training

The relation between the augmented pilot stanine of airplane commanders and the number of training requirements completed at the end of 6 weeks is illustrated in tables 4.31 and 4.32. In table 4.32 are shown the separate coefficients representing the correlation between augmented pilot stanines of B-17 and B-24 airplane commanders of 24 CCTS classes at 10 stations in the Second Air Force and the total number of all types of training requirements completed at the end of sixth week of training. In table 4.32 are given the correlations between stanines of the same airplane commanders and mission requirements completed in each training phase and the AAF Training Standards met. Also given in table 4.32 are similar data for B-29 airplane commanders at four stations.

TABLE 4.31.—*Correlations between stanines of airplane commanders and training accomplishment*

B-17 AND B-24 CREWS—SECOND AIR FORCE

Pilot stanine (with cr.) v. number of missions and requirements completed

Station	Number of cases	Requirements, 1 Mean	Requirements, 1 S. D.	Stanine, 2 Mean	Stanine, 2 S. D.	r_{12}
Alexandria, 10-27	41	5.88	1.66	82.0	5.0	-.18
Alexandria, 11-22	52	5.40	1.54	79.3	6.0	*.30
Alexandria, 1-20	23	6.04	1.65	94.2	3.8	.38
Alexandria, 2-18	8	5.50	1.50	80.5	4.3	.38
Bigs, 11-30	53	6.13	1.64	76.3	5.4	*.38
Bigs, 12-2	36	5.27	1.60	87.2	7.8	.38
Boise, 9-21	47	6.66	1.60	96.7	6.9	.30
Boise, 12-25	37	6.51	1.49	101.3	6.7	*.38
Boise, 2-11	42	6.23	1.74	91.4	6.8	.30
Boise, 3-10	32	6.13	1.66	81.2	7.6	.38
Casper, 11-30	47	6.15	1.40	84.3	5.6	.36
Casper, 12-30	46	6.63	1.72	85.5	6.6	.27
Dyersburg, 8-30	48	6.43	1.53	93.5	5.3	.28
Dyersburg, 9-22	50	6.30	1.66	91.9	6.5	.12
Mountain Home, 11-30	50	6.20	1.70	90.9	4.7	.08
Mountain Home, 12-11	45	6.27	1.71	85.9	5.7	*.38
Pueblo, 11-25	43	6.21	1.51	95.8	5.2	**.06
Pueblo, 12-25	38	6.27	1.57	107.6	4.0	.30
Rapid City, 9-28	47	6.23	1.62	93.2	6.3	*.27
Rapid City, 12-3	43	5.26	1.65	93.5	4.6	.36
Sioux City 11-20	43	5.86	1.65	76.6	6.0	-.17
Sioux City, 12-23	42	6.50	1.48	87.6	6.1	.38
Tucson, 12-17	44	6.48	1.57	86.7	6.9	**.06
Tucson, 11-20	43	6.16	1.67	86.4	6.6	.17
All stations (average by Fisher's z)	974	......	......	......	......	**.23

TABLE 4.32.—*Correlations between stanines of airplane commanders and training accomplishment*

SIXTH WEEK OF TRAINING—SECOND AIR FORCE

Type of score	N	Training Accomplished, 1 Mean	Training Accomplished, 1 S. D.	Pilot+Cr. stanine, 2 Mean	Pilot+Cr. stanine, 2 S. D.	r_{12}
B-29 airplane commanders:						
Mission requirements completed	161	35.3	8.90	6.60	1.59	0.08
AAF training standards met	155	36.1	17.4	6.59	1.62	.08
Number of hours flown	72	11.6	2.62	6.33	1.60	-*.26
B-17 and B-24 airplane commanders:						
Phase I, requirements	974	(1)	(1)	(1)	(1)	*.08
Phase II, requirements	974	(1)	(1)	(1)	(1)	*.38
Phase III, requirements	974	(1)	(1)	(1)	(1)	**.28
AAF training, standards met	974	(1)	(1)	(1)	(1)	**.15
All requirements	974	(1)	(1)	(1)	(1)	**.23

[1] Data are not available.

The data show that for B–17 and B–24 training the pilot stanine of the airplane commander did predict to some extent the relative number of training requirements that would be completed by his crew, at least in comparison with other crews in his class. The validity coefficient for the whole group of 974 airplane commanders was obtained by averaging by Fisher's z all coefficients for the separate classes. This procedure is particularly important in dealing with data of the sort indicated in table 4.31. The average number of training requirements completed varied considerably from class to class. Had this variance been included in the determination of the predictive value of the stanine, it is doubtful if the resulting coefficient would have differed significantly from zero.

The data obtained with B–29 airplane commanders do not support the findings with B–17 and B–24 training. However, it is of considerable interest to find that the pilot stanine had a negative relationship with number of hours flown. The implication of this result is that there would be a positive relationship between the pilot stanine and number of requirements completed per hour flown. Unfortunately, similar data were not obtained with the B–17 and B–24 crews.

Bombardment Scores

Although instructors were of the opinion that the proficiency of the airplane commander was reflected in part in the bombing scores attained by a crew, there was no evidence that the stanines of airplane commanders predicted such scores. The relevant data are presented in table 4.33. For the biserial coefficients in the latter half of the table, the 137 B–24 airplane commanders involved were divided into 2 groups on the basis of the size of the average circular errors of the crews. Biserial coefficients of validity were then computed. For the B–29 data, the small number of cases makes the negative coefficient obtained with radar bombing T-scores of doubtful significance.

TABLE 4.33.—*Correlations between stanines of airplane commanders and crew bombing scores*

SECOND AND FOURTH AIR FORCES

Air force	Aircraft	Type of score	Stanine	N	Bombing score, I: Mean	Bombing score, I: S. D.	Stanine, Z: Mean	Stanine, Z: S. D.	r_{IZ}
2d AF	B-17, B-24	Average circular error	P+	134	239	38	6.31	1.9	0.03
			P+	239	......	......	......	......	−.01
2d AF	B-29	Radar bombing T-score	P+	33	50.5	4.32	6.55	1.54	−.31
			N	33	50.5	1.22	5.79	1.45	−.03
				N_t	N_u	M_t	M_u	SD_t	r_{bis}
4th AF	B-24	Average circular error	B	137	60	6.21	6.20	1.73	0.00
			N	137	60	6.12	6.25	1.79	−.08
			P	136	60	6.20	6.13	1.49	.06

Performance in Ground Training

Table 4.34 presents coefficients of correlation between the augmented pilot stanines of B–17 and B–24 airplane commanders and various measures of performance in ground training. The coefficients were in general positive and several were significantly different from zero. In addition to the data in the table, ground school grade designations into upper and lower half of class were obtained for 160 B–29 airplane commanders in the 16th Wing in the Second Air Force. Stanines were available for 73 of these. The biserial coefficient of correlation between pilot stanine and ground school grades was 0.14 for this group.

Flying Evaluation Board Reports

A study was made in the Second Air Force of the pilots removed from flying status by action of the Central Flying Evaluation Board during a 2-year period. According to the available records, 257 pilots were so eliminated from operational training during this period, over 100 of these being eliminated in the last 6 months of the period. Separate figures were not available for fighter and bomber pilots since in many of the early cases such information was not recorded. Thus the data about to be reported refer to both bomber and fighter pilots.

Stanines were obtained for 203 of the 257 pilots removed from flying status by Board action. Since a total trainee population was not available

TABLE 4.34.—*Correlations between stanines of airplane commanders and ground training criteria*

B-17 AND B-24 AIRPLANE COMMANDERS—SECOND AIR FORCE

Type of score	Number of cases	Stanine, 1		Criterion, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Academic average	189	6.54	1.6	85.6	3.2	*0.14
Ground school grades (standard scores)	188	6.14	1.6	54.1	10.6	*.16
Engineering grade	35	6.34	1.7	85.3	5.1	.18
Gunnery grade	37	6.38	1.7	82.0	4.6	.25
Bombardment grade	37	6.38	1.7	85.8	7.0	−.08
Communications grade	37	6.38	1.7	96.9	1.3	.01
Link trainer grade	49	6.20	1.6	17.8	2.8	.30
Hours to complete link training	104	6.26	1.6	12.3	2.6	*.15
C-1 autopilot examination	48	6.18	1.6	86.9	9.1	*.33
	Type of stanine	N_s	P_c	M_c	M_i	SD_s	r_{bis}
Completion of link trainer check	P+	83	0.67	6.11	6.07	1.63	0.08
	N	83	.67	6.73	5.93	1.72	*.29

c = Completed Link Trainer check.
i = Did not complete Link Trainer check.

from which to obtain the necessary statistics, an attempt was made to construct a control group for validation of the stanine against the criterion of elimination from operational training by Board action. The control group was obtained by locating each eliminated pilot in the Second Air Force Command rosters, and then taking down the names of the nearest four pilots in the roster who had the same duty specialty listing. For eliminated pilots without duty specialty listing, the nearest four pilots were chosen without regard for duty specialty. The four pilots chosen in each case were then looked up in the AAF Training Command rosters of stanines of pilots graduated from advanced pilot schools. In some cases, stanines were not available for any of the four names. Where stanines were available for more than one of the four, the one chosen was the pilot whose date of testing was closest to that of the eliminated pilot. For pilots tested prior to the introduction of mark-sensing procedures, it was also necessary to control the Research Unit at which the testing occurred. Controls were obtained in this manner for 1[illegible]4 of the eliminated pilots.

In table 4.35 below are given the average stanines for the pilots removed from flying status by Flying Evaluation Board Action, and for the control pilots in training at the time of such elimination. A small and not significant difference was found between the stanines of the two groups.

Because of some serious sources of error in the data, it was difficult to evaluate the results. Some sources of error tended to minimize any differences between the two groups, others tended to increase such apparent differences. For example, the average stanines of pilots entering operational training probably increased over the 2-year period of the study. This is corroborated by the fact that the average stanine of the control pilots in table 4.36 is slightly higher than that of the control pilots in table 4.35. The

TABLE 4.35.—*Comparison of stanines of pilots removed from flying by board action and those of control groups*

SECOND AND THIRD AIR FORCES

Second Air Force

Type of stanine	N	Pilots removed by board action		Control pilots		P_{diff}
		Mean	S. D.	Mean	S. D.	
Pilot+credit	184	5.83	1.82	6.05	1.89	0.14
Pilot + credit	[1]28	5.04	1.41	5.72	1.65	.06

Third Air Force

Type of stanine	N	Pilots removed by board action Mean	Pilots removed by board action S. D.	Control pilots Mean	Control pilots S. D.	P_{diff}
Bombardier	40	5.65	1.78	6.39	1.33	0.06
Navigator	40	5.41	2.20	6.26	1.86	.02
Pilot	40	5.48	1.70	6.37	1.73	<.01

[1] Only pilots tested prior to mark-sensing records procedures. This group is included in the 184 cases given above it.

pilots in 4.36 were in general those most recently in training since the eliminated pilots with which they were paired were those that had relatively more detailed and complete records of Board action. Thus if pilots removed from flying status by Flying Evaluation Board action had been held in the air force for 12 or 18 months awaiting Board action—as frequently happened—controls chosen on the basis of time of Board action might have belonged to a population with a different average stanine.

On the other hand, it was conceivable that controls tested at the same time and remaining in the air force in operational training until the same time, while most of their classmates had long since been sent overseas, did so because they also were doing poorly in such training and were forced to repeat the training in later classes. From interviews with training personnel, it was gathered that for a long time officers who failed in operational training were either transferred to other duties without Flying Evaluation Board action or else were reassigned to a later class for further training. This was contrary to most recent practice, but it may have tended to minimize differences between eliminees and controls chosen in the way they were chosen.

The reasons for removal from flying status by Flying Evaluation Board

action can be divided roughly into three classes: medical, unsatisfactory performance and LIF (lack of intestinal fortitude) or fear. In table 4.36 below are given the average stanines of pilots eliminated for these three classes of reasons, together with the averages for the corresponding control groups. Such data were available for only 101 of the 184 pilots removed by Board action.

TABLE 4.36.—Comparison of pilot stanines of pilots removed from flying by board action and those of a control group

SECOND AIR FORCE

Reason for FEB action	N	Eliminated pilots		Control pilots		P
		Mean	S. D.	Mean	S. D.	
Fear and lack of intestinal fortitude	28	6.32	1.86	6.14	1.69	−0.35
Unsatisfactory performance	32	6.09	1.97	6.19	1.75	.35
Physical reasons	41	6.22	1.63	6.20	1.99	−.48

A somewhat similar project was carried out in the Third Air Force where a total of 101 aircrew officers were removed from operational training for reasons other than medical during 1944. In this group were 68 pilots, 17 bombardiers and 16 navigators. It represented approximately 0.4 percent of the total number of such officers trained in the Third Air Force in 1944. Of the 68 pilots removed from training approximately half were airplane commanders and the other half were fighter pilots. Because of the small numbers involved no attempt was made to separate the pilots according to type.

The stanines of the pilots for whom they were available were compared with stanines of an equal number of pilots chosen as controls. An attempt was made to match each airplane commander removed from flying status with an officer not so removed who had been in training in the same CCTS class and who had been tested at the same Psychological Research Unit and who was in the same AAF Training Command class. It was not possible to match the groups in terms of such a complete set of factors. Therefore, the control cases were matched solely in terms of Psychological Research Unit at which they were tested and AAF Training Command class. The means and standard deviations of the stanines of both groups are compared in table 4.33 above. All comparisions reveal that the pilots removed had lower stanines than those of the control group.

Aircraft Accidents

Studies were made in all of the Continental Air Forces of the validity of stanines of pilots in predicting occurrence of aircraft accidents in operational training. Data for aircraft accidents of fighter pilots have already been reported in Chapter II, Fighter Pilot. Data on the aircraft accidents of bombardment pilots have been presented in an earlier section of this

chapter under Criteria. The relation between the stanines of the airplane commanders concerned and the occurrence of aircraft accidents will be discussed in the paragraphs which follow.

The four-engine aircraft accidents occurring in a 6-month period in the Second Air Force were analyzed by research personnel in that command. During the period concerned 396 accidents were reported for four-engine aircraft. Stanines were obtained for 174 of the airplane commanders involved in these accidents. While this group included 62 percent of the B-17 and B-24 pilots, only eight percent of the B-29 pilots involved in these accidents were included. The results of the study were, therefore, not representative of B-29 accidents.

Since neither a total training population nor the statistical characteristics of such a population were immediately available for use in comparing stanines of various special groups, it was necessary to construct a pilot population to serve as a control for comparing their stanines with those of the accident pilots. The first attempt to obtain a control group involved locating the accident pilots on the AAF Training Command rosters of pilots graduating from Advanced Pilot School. A control group was then obtained by selecting the name of the pilot in the roster immediately following that of the accident pilot. Where that pilot also had an accident, the name of the immediately preceding pilot was substituted. This control group was designated Control Group A.

A second control group was obtained in a somewhat different manner. Here the pilot having an aircraft accident was located on the Command Rosters of the Second Air Force. The names of the two preceding and the two following pilots with the same duty status and assignment were copied down. These names were then looked up in the Rosters of Stanines. If stanines were available for more than one of these pilots, the one chosen was the pilot whose testing number was closest to that of the pilot having the accident. This control group was designated Control Group B.

In table 4.37 are shown the stanine means and standard deviations of the pilots having accidents due to different causes together with similar data for Control Groups A and B.

In the Third Air Force studies of aircraft accidents, research personnel made separate analyses of stanine validity in predicting accidents in medium and in heavy bombardment aircraft. As in the Second Air Force a control group of pilots was chosen to make comparisons of stanines. First, the controls were so chosen that they were pilots in training in the Third Air Force at the time of the accident. Then, further, the controls were pilots that had been examined in the same Psychological Research Unit and had been given the same battery of classification tests as the pilots with accidents. The average stanines of the Third Air Force pilots suffering aircraft accidents and those of the controls are also given in table 4.37 above, together with the corresponding standard deviations. Separate averages are given for the pilots having the 40 medium bombardment air-

craft accidents and for those having the 122 heavy bombardment aircraft accidents.

A similar study was made in the Fourth Air Force. It included all B-24 aircraft accidents occurring between 1 January 1944 and 1 March 1945. On the basis of the reports of Aircraft Accident Investigation Boards the accidents were separated into those reported as caused primarily by pilot error and those of unknown cause or causes beyond the control of the pilot. Out of a total of 161 airplane commanders involved in one or more air-

TABLE 4.37.—*Comparison of stanines of airplane commanders having aircraft accidents with those of control groups without accidents*

SECOND, THIRD AND FOURTH AIR FORCES

Second Air Force

Origin of accident	Number of cases	Accident pilots		Control group A		Control group B	
		Mean	S. D.	Mean	S. D.	Mean	S. D.
Pilot error	51	6.11	1.54	6.22	[illegible].53	[illegible].45	1.[illegible]
Unknown	36	6.34	1.71	6.29	1.49	[illegible].67	1.[illegible]
Materiel and unavoidable	87	6.39	1.58	6.23	1.59	[illegible]	1.[illegible]
Total	174	......	......	......	......	......	....

Third Air Force

Aircraft	Type of stanine	Number of cases	Accident pilots		Number of cases	Control pilots		Pag
			Mean	S. D.		Mean	S. D.	
Medium bombardment	Bombardier	40	5.93	1.54	40	6.30	1.50	[illegible]
	Navigator	40	5.75	1.74	40	6.[illegible]	1.55	.[illegible]
	Pilot	40	5.80	1.40	40	6.23	1.54	.[illegible]
Heavy bombardment	Bombardier	122	6.0	1.9	154	6.0	1.6	.[illegible]
	Navigator	122	6.0	1.8	154	6.0	1.7	[illegible]
	Pilot	122	6.1	1.7	154	6.0	1.6	−.[illegible]

Fourth Air Force

Origin of accident	Number of cases	Accident pilots		Control pilots		Pag
		Mean	S. D.	Mean	S. D.	
Pilot error	81	6.05	1.65	......	......	....
Unknown	11	5.91	1.44	......	......	....
Not pilot error	56	6.19	1.62	......	......	....
All accidents	148	6.09	1.64	6.13	1.60	0.[illegible]

craft accidents, 148 were found to have been given the Aircrew Classification Tests so that stanines were available for this number.

A control group of approximately 350 airplane commanders who had had no aircraft accidents as far as was known were selected from the population of combat crews trained during the period of the study. The numbers selected from crews in training in each quarter of the year were proportional to the number of accidents for the quarter. Of the 350 control airplane commanders selected, stanines were found for 302. The average pilot stanines of pilots having accidents due to pilot error and to other causes and of the control pilots are shown in table 4.37 already presented.

The studies in all three air forces in table 4.37 show that in general the stanines of control groups of pilots were higher than those of airplane commanders having aircraft accidents. Although the differences are in the expected direction, with one exception they are not individually sufficiently large to negate the hypotheses they are chance fluctuations. However, collectively they indicate a small positive correlation.

Selection as Airplane Commander or Copilot

The average pilot stanines of airplane commanders and copilots in the Second and Fourth Air Forces are shown in table 4.38. There is obviously a marked difference between the stanines of airplane commanders and those of copilots. In the first comparison in the Second Air Force data, where the difference was not so marked, it is interesting to report that at the stations at which these pilots were assigned in the Second Air Force there was great dissatisfaction with the assignments made. In several cases observed directly by research personnel, copilots and airplane commanders were reversed in their assignment by their directors of Training. Further inquiry revealed that many of the copilots in operational training at this time had been two-engine instructor pilots in the AAF Training Command. Orders from higher headquarters to assign these pilots to combat outfits caused them to be sent on to operational training where they automatically became copilots since they had had no 4-engine experience.

TABLE 4.38.—*Comparisons of stanines of airplane commanders and copilots*

B-17 AND B-24 CREWS—SECOND AND FOURTH AIR FORCES

Air force	Duty assignment	N	Pilot stanine Mean	Pilot stanine S. D.
Second Air Force, first half—1944	Airplane commander	329	6.46	1.75
	Copilot	329	6.03	1.51
Second half—1944	Airplane commander	974	6.24	1.66
	Copilot	803	5.39	1.63
Fourth Air Force, second half—1944	Airplane commander	1,218	6.18	.....
	Copilot	1,232	5.14	.....

In utilizing this comparison as a criterion of proficiency it is presumed that graduates of advanced flying schools were assigned to duty as airplane commanders or copilots on the basis of the degree of competence demonstrated to date. In stanine validation it is also assumed that the assignment was done without knowledge of the stanines of the pilots involved. It is believed from reports of procedures used at advanced schools in the AAF Training Command that the pilot stanine was not often utilized directly in making the decision as to assignment. However, since the stanine (or related information) did enter into the decision as to whether the individual was commissioned as a Flight Officer or Second Lieutenant, the criterion is open to objection when used to validate the stanine.

Test Validity

The validities of individual Classification Test scores were computed for a small sample of B-24 airplane commanders in the Fourth Air Force. Criteria used were Instructors' Ratings, Link Trainer Grades and Crew Average Circular Error. All criterion scores were converted to a two-category score by dividing them into two approximately equal groups of upper and lower half scores. Validity coefficients were then computed in biserial form. The data are shown in table 4.39-4.41.

Only a few of the individual classification tests had validity coefficients that differed significantly from zero. These were Mechanical Principles, CI903A; Biographical Data (P) and (N); CE602D, and Spatial Orientation I, CP501B.

TABLE 4.39.—*Validities of stanines and classification test scores for instructors' ratings of airplane commander (B-24) proficiency*

BATTERIES 1, 2, 3, 4

Test variables	N_t	N_u	M_t	M_u	SD_t	r_{bis}
Bombardier stanine	130	59	6.18	5.68	1.71	**0.33
Navigator stanine	130	59	6.06	5.64	1.76	*.27
Pilot stanine	129	58	6.20	5.83	1.53	*.27
Rotary pursuit w/DA CP410B	106	64	51.51	51.70	9.35	-.08
Two hand coordination CM101A	128	58	52.68	51.64	10.30	.11
Complex coordination CM701A	130	59	52.16	52.92	9.69	-.09
Aiming stress CE211A	124	56	50.31	48.91	11.53	.14
Discrimination reaction time CP611D	130	59	53.23	52.22	7.62	.15
Finger dexterity CM116A	130	59	50.12	48.97	11.10	.12
Reading comprehension CI614G	124	56	24.78	23.68	10.94	.11
Mechanical principles CI903A	124	56	63.27	57.80	18.01	**.35
Dial and table reading CP622-21A	130	59	36.15	35.20	7.82	.14
Spatial orientation IICP503B	130	59	22.19	21.51	6.82	.11
Spatial orientation ICP503B	130	59	29.18	29.80	5.66	-.12
Numerical operations ICI702B	127	59	37.67	36.15	10.95	.16
Numerical operations IICI702B	127	59	34.04	32.29	11.21	.18
Biographical data (P) CE602D	95	46	28.58	27.76	6.48	.16
Biographical data (N) CE602D	95	46	23.01	22.28	3.16	*.21
Speed of identification CP610A	127	49	33.20	32.25	7.59	.15
General information (N) CE505D	95	46	21.36	21.78	5.56	-.09
General information (P) CE505D	95	46	34.97	35.15	6.12	-.04
Mathematics B CI206C	95	46	16.96	16.33	9.03	.08
Mathematics A CI702F	95	46	27.08	25.48	16.48	.12

TABLE 4.40.—*Validities of stanines and classification test scores for link trainer grades made by airplane commanders (B-24)*

BATTERIES 1, 2, 3, 4

Test variables	N_t	N_u	M_t	M_u	SD_t	r_{bis}
Bombardier stanine	137	82	6.18	6.07	1.73	0.07
Navigator stanine	137	82	6.01	5.94	2.02	.05
Pilot stanine	138	84	6.14	6.00	1.54	.14
Rotary pursuit w/DA CP410B	97	45	52.05	51.73	9.35	.04
Two hand coordination CM101A	135	82	52.54	51.55	10.37	.15
Complex coordination CM701A	137	82	52.02	53.10	9.63	-.16
Aiming stress CE211A	131	80	50.45	50.09	11.53	.05
Discrimination reaction time CP611D	137	82	53.47	53.85	7.78	-.08
Finger dexterity CM116A	137	82	49.93	51.20	10.[illegible]	-.18
Reading comprehension CI614G	131	80	24.68	23.34	11.10	.19
Mechanical principles CI903A	131	80	62.63	60.56	17.82	.18
Dial and table reading CP622-21A	137	82	36.39	36.01	6.81	.08
Spatial orientation II CP503B	137	82	22.31	22.07	6.88	.05
Spatial orientation I CP503B	137	82	29.24	28.84	6.23	.10
Numerical operations I CI702B	134	81	38.28	38.02	11.38	.04
Numerical operations II CI702B	134	81	34.63	34.67	1[illegible].03	.09
Biographical data (P) CE602D	104	63	28.63	27.60	6.44	*.23
Biographical data (N) CE602D	104	63	23.12	22.87	3.24	.12
Speed of identification CP610A	134	81	33.47	33.64	7.[illegible]	-.03
General information (N) CE505D	104	63	21.[illegible]	20.92	5.89	.11
General information (P) CE505D	104	63	34.78	34.83	6.51	-.01
Mathematics B CI206	104	63	16.88	16.14	9.72	.13
Mathematics A CI702F	104	63	27.61	26.21	16.85	.13

TABLE 4.41.—*Validities of airplane commanders' (B-24) stanines and classification test scores for crew average circular bombing error*

BATTERIES 1, 2, 3, 4

Test variables	N_t	N_b	M_t	M_b	SD_t	r_{bis}
Bombardier stanine	137	60	6.21	6.20	[illegible]	.00
Navigator stanine	137	60	6.12	6.25	[illegible]	−.08
Pilot stanine	136	60	6.20	6.13	1.49	.06
Rotary pursuit w/DA CP410B	102	46	51.55	50.33	9.32	.15
Two hand coordination CM101A	135	58	52.75	52.43	10.16	.03
Complex coordination CM701A	137	60	52.43	52.47	9.91	−.08
Aiming stress CE211A	131	57	50.47	49.86	11.68	.06
Discrimination reaction time CP611D	137	60	53.39	53.80	7.74	−.06
Finger dexterity CM116A	137	60	50.23	49.02	11.09	.12
Reading comprehension CI614G	131	57	25.29	26.37	11.73	−.10
Mechanical principles CI903A	131	57	62.56	61.91	17.99	.04
Dial and table reading CP622-21A	137	60	36.46	36.80	6.92	−.06
Spatial orientation II CP503B	135	60	22.62	22.43	6.35	.03
Spatial orientation I CP401B	135	60	29.99	28.98	5.77	*.20
Numerical operations I CI702B	134	58	38.23	37.38	11.37	.08
Numerical operations IICI702B	134	58	31.25	[illegible]	11.02	−.01
Biographical data (P) CE602D	100	45	28.45	29.20	6.23	−.14
Biographical data (N) CE602D	100	45	23.18	33.40	3.22	−.08
Speed of identification CP610A	134	58	33.14	33.72	7.83	−.08
General information (N) CE505D	100	45	21.46	22.20	5.96	−.16
General information (P) CE505D	100	45	34.90	33.91	6.49	.17
Mathematics B CI206C	100	45	18.73	16.20	10.13	.06
Mathematics A CI702F	100	45	27.20	30.62	16.58	.30

SUMMARY

Aviation psychologists in the Continental Air Forces spent a relatively large proportion of their time and energy in the study of airplane commander proficiency in operational training. Conferences were held with training personnel, combat crews in training, and personnel experienced in combat. Intelligence reports, Training Manuals and Directives and various training forms were examined and studied. These activities not only gave the research personnel information as to the types of characteristics needed for success in operational training, but they also provided the raw material for the construction of several instruments designed to measure proficiency in that training.

A large number of possible criteria of proficiency were analyzed, most of which turned out to be unacceptable for one reason or another. Because of the lack of suitable objective criteria, aviation psychologists were forced to fall back on ratings and other subjective judgments. Among the subjective criteria studied were over-all numerical ratings of various sorts, descriptive ratings, rankings, mission ratings, check list ratings and self ratings. Other criteria included rate of training accomplishment, bombardment scores (both average circular error and radar bombing scores), academic grades, ground trainer scores, Flying Evaluation Board reports, and aircraft accidents.

Wherever data were available, study was made of the validity of the stanines of airplane commanders in predicting proficiency in operational training. It was found that pilots removed from flying status by Board action and pilots having aircraft accidents as a group had lower stanines than control groups of other pilots in training at the same time. No relationship was found between pilot stanine and bombardment scores. Whether

this lack of relationship was due to various attenuating factors such as the unreliability of the bombing scores and the restriction of range of the stanine is not known. Certainly the data did not support the contention of instructors that the pilot contributed in any important way to the accuracy of bombing. Perhaps beyond a certain rather low level of performance the skill of the pilot did not make much difference in bombing.

The pilot stanine did predict to some extent the more adequate types of ratings. However, the coefficients were all low. Somewhat higher predictive values were obtained in the prediction of training accomplishment scores and ground training performance. Among the ratings, the most objective and in many ways the most interesting were the ratings on the Crew Proficiency Check—VHB. Unfortunately only fragmentary data were obtained and no conclusions regarding their use are possible.

CHAPTER FIVE

Copilot

ANALYSIS OF DUTIES

Job Description

The copilot on heavy and very heavy bombardment crews served as an assistant or understudy of the airplane commander. In this capacity he was supposed to be familiar with every duty of the latter, since he had to be able to take the place of the airplane commander at any time. On long combat and training flights the copilot and airplane commander usually took turns flying the plane. On landings and take-offs the copilot assisted by adjusting flaps, raising and lowering wheels, checking instruments and performing various operations at the signal of the airplane commander. In general, in operational training the copilot received approximately one-half as much "stick" time as the airplane commander.

Other duties of the copilot varied considerably with the type of aircraft and the specific division of duties worked out between particular airplane commanders and their copilots. The copilot usually performed all or part of the preflight and postflight checks, usually under the supervision of the airplane commander. In heavy bombardment crews the copilot frequently served as engineering officer and maintained a complete log of power plant performance. He thus had to have a thorough knowledge of cruise control and emergency procedures. In many cases the copilot handled routine dealings with other crew members for the airplane commander and accomplished much of the paper work with respect to training and administration. The tendency grew up, particularly in some of the combat areas and in B–29 training and combat operations, for the more experienced and older pilots to be assigned as airplane commanders. The position of copilot then served as a sort of apprenticeship from which a pilot graduated to airplane commander status after a given amount of experience in combat.

The actual training of the copilot was very similar to that of the airplane commander. He took the same classes in ground training, flew when the airplane commander flew and had few special duties or opportunities to come to the attention of training supervisors. The copilot thus came to be the least well known of the aircrew officers. Flying instructors sometimes did not even know who were the copilots of the crews under their instruction. For a more detailed statement of the training requirements, see the material under the corresponding section in chapter 4, Airplane Commander.

Job Specifications

Aviation psychologists in the Second, Third and Fourth Air Forces conferred with training personnel, officers with combat experience, and combat crew trainees to obtain information about qualities and skills necessary for success as copilot on bombardment crews. There was general agreement that the primary requisite was a high degree of proficiency in handling the aircraft. Of especial importance were proficiency in formation and in instrument flying. These same qualities were also those thought most important for success in lead crew assignment, according to opinions of Second Air Force personnel interviewed about lead crew proficiency. Apparently there were few or no differential qualities or traits for copilots as distinct from those required for success as airplane commander.

CRITERIA OF PROFICIENCY

The studies of criteria of proficiency of copilots parallelled those for airplane commander. Hence the various criteria will be described here only in sufficient detail to provide accurate identification. The full descriptions may in each case be found in chapter 4, Airplane Commander, under the section on Criteria. In this connection it should be pointed out that the amount of information available about copilots was much less than was available for airplane commanders. For example, it was sometimes impossible to obtain subjective ratings of copilots in CCTS classes in operational training because none of the instructors or supervisory personnel felt they knew the copilots well enough to make the ratings.

Ratings of Proficiency

Over-all Judgments

Efficiency ratings on WD AGO Form 66-2, 5-point ratings of attitude toward work, ratings on the Officers Proficiency Card, AFTRC Form 2 and rankings in flights and classes were obtained on small samples of copilots. No data are available on the reliability of any of these ratings as applied to copilots.

Descriptive Ratings

In developing procedures for measuring lead crew proficiency of combat crews in operational training, two sets of descriptive rating scales, Scale A and Scale B, were constructed by research personnel in the Second Air Force. Included were scales for rating the crew as a whole, scales for rating enlisted gunners as a group, and scales for rating each aircrew specialty. Details of the construction of these scales are given in the section on Criteria in chapter 4, Airplane Commander. In both Scale A and Scale B the items included for the rating of copilots were similar to those in the airplane commander scales which were described in chapter 4. Scale A was used to rate 160 B–29 copilots at 4 stations and Scale B was used to

rate between 150 and 200 additional copilots at another 4 stations. Each copilot was rated independently by two or more instructors. The distributions of the totals of the ratings on all items are given in table 5.1 for both scales. The correlations between the ratings of two instructors are given in table 5.2. The data show that the reliability of copilots' ratings was only moderate and was somewhat lower than that of ratings of airplane commanders. A lower reliability for the ratings of copilots is not unexpected in view of the fact that copilots were much less well known to the raters.

TABLE 5.1.—*Distribution of descriptive ratings of copilots*

SCALES A AND B—B-29 CREWS—SECOND AIR FORCE

Scale A-5 Items Rating:	N	*Scale B-7 Items* Rating:	N
5	3	6–7	17
6	6	8–9	10
7	8	10–11	16
8	29	12–13	22
9	39	14–15	56
10	45	16–17	25
11	54	18–19	49
12	75	20–21	70
13	34	22–23	31
14	19	24–25	15
15	5	26–27	1
16	2	28–29	5
17	1	30–31	1
		32–33	1
Total rating	320	Total rating	319
Mean	10.83	Mean	16.62
S.D.	2.08	S.D.	5.14

TABLE 5.2.—*Reliability of ratings of copilots*

B-29 CREWS—SECOND AIR FORCE

Station	N	First rater, 1		Second rater, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Scale A:						
Alamogordo 2	23	1.90	0.28	2.01	0.29	0.43
Alamogordo 6	23	1.95	.40	1.92	.42	.74
Biggs Field	41	2.42	.34	1.96	.38	.66
Clovis	36	2.31	.34	2.27	.37	.57
Tucson	37	2.05	.37	2.55	.30	−.08
All stations (by Fisher's z)	160	2.16	.41	2.17	.44	.49
Scale B:						
Albuquerque	26	2.16	.42	2.21	.4[illegible]	.08
Great Bend	13	1.87	.82	2.16	.78	.83
Pyote	46	2.77	.44	2.92	.59	.51
Walker	42	2.49	.85	2.04	.78	.65
All stations (by Fisher's z)	127	2.47	.71	2.40	.75	.54

The aircrew officers in one B–24 CCTS class in the Second Air Force were interviewed by an Interview Board which then rated each copilot on a 10-point numerical scale on each of 6 traits. Further details of these ratings may be found in chapter 4, Airplane Commander, under Criteria.

The correlations between the Interview Board Rating on the different traits are given in table 5.3. The negative relationships found between the trait ratings of the airplane commanders were not duplicated here with the copilots.

TABLE 5.3.—*Correlations between interview board ratings of six traits*

B-24 COPILOTS—10-POINT SCALE—SECOND AIR FORCE[1]

Trait	Code	A	B	C	D	E	F
Interest in flying	A	......	0.27	0.42	0.20	0.03	0.25
Flying experience	B	......	......	.09	.14	.33	.38
Desire for combat	C	......	......	......	.19	.03	.30
Getting along with others	D	......	......	......	......	.43	.70
Educational background	E	......	......	......	......	......	.44
Officer qualities	F	......	......	......	......	......	......

[1] N = 54.

Performance in Ground Training

A number of proficiency measures in ground training were studied by aviation psychologists in the Continental Air Forces. Types of scores studied with B-17 and B-24 copilots were academic average, grades in Engineering, Bombardment, Communications, Navigation and Gunnery, Link Trainer grades or ratings, hours taken to complete Link Trainer check, circular error on the A-2 Bomb Trainer and examination scores in C-1 Autopilot and Engineering. A small amount of evidence was available on the reliability of academic grades. The correlation between Engineering grade in the first half and that in the second half of training was 0.37 for 53 B-17 copilots. The correlation between academic average for all courses in the first half of training and that for the total course was 0.86 for 56 B-24 copilots. The latter figure is spuriously high since the performance of the first half was probably included in the grade for the total course.

For 56 copilots at one station in the Fourth Air Force, the odd-even reliability of Link Trainer ratings for 5 missions was 0.46.

Miscellaneous Criteria

Crew rate of accomplishment of training requirements, crew bombing scores, and removal from flying status by action of Flying Evaluation Boards were also studied as criteria of copilot proficiency. However, there was little reason to expect that the copilot contributed significantly to crew performance in the first two types of measures. And in the Flying Evaluation Board reports, few of the re-evaluated pilots were reported to be copilots. Whether this was because the reports frequently did not contain the necessary information or whether the copilot actually was rarely involved in Board action was not certain. Reports of instructors and others would indicate that the former was the more likely explanation.

Intercorrelations Among Criteria

The correlations between various criteria of proficiency of copilots are shown in tables 5.4 and 5.5. Correlations between various types of ratings

in the Second Air Force are given in the first of the two tables. The second table gives the intercorrelations among certain ground training proficiency criteria and an over-all rating by instructors. In both tables coefficients of correlation are given above and to the right of the diagonal and numbers of cases for those coefficients below and to the left.

The correlations between descriptive ratings and certain other proficiency criteria were determined for various groups of B–29 copilots. Ratings of

TABLE 5.4.—*Correlations between different ratings of copilots*

B–17 AND B–24 COPILOTS—SECOND AIR FORCE

Type of rating	Code	Correlations between ratings: A	B	C	D	E
Efficiency rating, form 66-2	A		0.37	0.09	0.04	0.33
Rank in class	B	167		.22	.20	.25
AFTRC Form 2:						
Military discipline	C	101	101		.90	.28
Manner of performing duties	D	101	101	101		.36
General aircrew ability	E	101	101	101	101	

GENERAL NOTE—Entries below the diagonal are the numbers of cases for the corresponding correlation coefficients.

160 B–29 copilots on Scale A in the Second Air Force were found to correlate 0.00 and 0.02 with academic average and average circular error. These coefficients were obtained by averaging the coefficients for the four different stations involved. For 39 copilots of B–29 crews in the Second Air Force the correlation between ratings on descriptive rating Scale B and

TABLE 5.5.—*Correlations between measures of proficiency of copilots*

B–24 CREWS—FOURTH AIR FORCE

Type of score	Code	Correlations between scores: A	B	C	D	E
Over-all rating	A		−0.16	−0.20	0.30	......
Link trainer rating	B	41		.17	.26	0.14
A-2 circular error	C	40	41		.22	......
C-1 autopilot examination	D	40	41	40		......
Engineering examination	E	......	102	......	......	

GENERAL NOTE—Entries below the diagonal are the numbers of cases for the corresponding correlation coefficients.

number of hours to complete Link Trainer requirements was 0.18. For another 44 copilots at another station, the biserial correlation between ratings on Scale B and whether or not the copilots completed their Link Trainer Check was 0.25.

Evaluation

An adequate evaluation of criteria of proficiency for copilots is difficult with the small amount of data available. Because they were so little known to instructors and training personnel there is good reason to doubt the validity of ratings. For example, for 8 CCTS classes of B–17 and B–24 crews, instructors ranked the aircrew officers in their flights in the order of their expected effectiveness in combat. The average (by Fisher's z) cor-

relation between the ranks given airplane commanders and those given copilots in the different CCTS classes was 0.68. It would seem rather obvious that a copilot was usually ranked in the class or flight according to what the instructors thought of the airplane commander.

Performance on ground trainers would seem to be the best of the available criteria for a number of reasons. The scores represented observations or records of actual performance and were thus not influenced by lack of knowledge of the individual in the way the over-all and descriptive ratings were. Similarly they were not affected as much by various forms of subjective bias since attention of the observers was focussed on actual performance. The chief disadvantage to ground trainer scores was that they did not necessarily represent the type of activity actually performed in the air.

VALIDATION

The correlations between the augmented pilot stanines of B-17 and B-24 copilots in the Second Air Force and various types of ratings of proficiency are shown in table 5.6. Except for the Interview Board ratings, none of the coefficients differ significantly from zero. The latter coefficient is based on such a small number of cases that its practical significance is open to question.

In table 5.7 are shown the correlations between the stanines of B-29 copilots and descriptive ratings in Scales A and B in the Second Air Force.

TABLE 5.6.—*Correlations between augmented pilot stanines and ratings*

B-17 AND B-24 COPILOTS—SECOND AIR FORCE

Criterion	Number of Cases	Stanines, *1*		Ratings, *2*		r_{12}
		Mean	S. D.	Mean	S. D.	
Efficiency rating, form 66-2	155	5.70	1.64	42.4	10.5	0.05
Rankings by instructors	130	5.84	1.63	50.7	18.2	.08
General aircrew ability, AFTRC Form 2	40	5.13	1.71	51.4	8.2	−.02
Attitude toward work	211	6.04	1.70	2.10	.3	.09
Interview board ratings	28	5.81	1.44	16.4	8.3	*.42

The coefficients shown are averages (by Fisher's z) of separate coefficients for each station and class. None are significantly different from zero.

Correlations between augmented pilot stanines of B-17 and B-24 copilots and various criteria of performance in ground training are shown in table 5.8. Only the coefficient for the engineering grades was significantly (< 5

TABLE 5.7.—*Correlations between stanines of copilots and ratings by instructors*

SCALES A AND B, B-29 CREWS—SECOND AIR FORCE

Scale	Stanine	N	Stanine, *1*		Rating, *2*		r_{12}
			Mean	S. D.	Mean	S. D.	
Scale A	Bombardier	156	6.14	1.84	2.17	0.35	−0.04
Scale B	Bombardier	172	6.31	1.66	2.58	.58	−.06
Scale A	Navigator	156	5.95	1.75	2.17	.35	.00
Scale B	Navigator	172	5.86	1.86	2.58	.58	.04
Scale A	Pilot + Cr.	156	6.67	1.54	2.17	.35	−.02
Scale B	Pilot + Cr.	172	6.42	1.51	2.58	.58	.06

percent) different from zero. For the CCTS class of 54 B–24 crews in the table above the navigator and bombardier stanines of the copilots were found to better predict navigation and bombardment grades than did the pilot stanine. The coefficients were 0.31 for bombardier stanine against Bombardment grade and 0.18 for navigator stanine against Navigator grade. Finally the navigator stanine was also found to better predict academic average. For the 211 copilots in the first line of table 5.8 the correlation between navigator stanine and academic average was .27. For

TABLE 5.8.—*Correlations between augmented pilot stanine and ground training criteria*

B–17 AND B–24 COPILOTS—SECOND AIR FORCE

Criterion	Number of Cases	Pilot stanine, 1: Mean	Pilot stanine, 1: S. D.	Training score, 2: Mean	Training score, 2: S. D.	r_{12}
Academic average	211	6.04	1.70	81.3	4.5	0.00
Ground school grades	160	5.52	1.63	50.9	15.0	.04
Engineering grade	54	5.48	1.54	81.4	5.4	*.24
Gunnery grade	54	5.48	1.54	83.3	5.3	.07
Bombardment grade	34	5.48	1.54	84.7	3.2	.19
Communications grade	54	5.48	1.54	97.3	1.2	.07
Navigation	54	5.48	1.54	78.7	8.6	−.09
Hours to complete link training	45	5.92	1.59	13.5	2.5	.06
C-1 autopilot examination	45	5.92	1.59	79.4	11.6	.11

157 B–29 copilots in the Second Air Force, upper and lower half designations were available for performance in ground school. The coefficients of correlation between augmented pilot stanines of these copilots and performance in ground school averaged (by Fisher's z) was —0.02.

In table 5.9 are shown the correlations between pilot and navigator stanines and performance on the Link Trainer. Two types of scores were available, hours to complete Link Trainer requirements and whether or not the Link Trainer check was completed. The latter score was, of course, a sort of pass-fail criterion and those coefficients were biserial in form. It was interesting that the navigator stanine seemed to predict to a greater extent the copilots' performance in the Link Trainer than did the pilot stanine.

In chapter 4, Airplane Commander, it was reported that records of more than 1,200 crews in 24 CCTS (H) classes were analyzed to determine the

TABLE 5.9.—*Correlations between stanines of copilots and link trainer performance*

B–29 CREWS—SECOND AIR FORCE

Type of stanine	N	Stanines, 1: Mean	Stanines, 1: S. D.	Link trainer score, 2: Mean	Link trainer score, 2: S. D.	r_{12}
Hours to complete link trainer requirements						
Pilot + Cr.	34	6.24	1.44	10.85	2.29	0.22
Navigator	34	6.26	2.05	10.85	2.29	*.33
	N_t	P_c	M_c	M_t	SD_t	r_{bis}
Completion of link trainer check						
Pilot + Cr.	46	0.63	6.18	6.10	1.33	0.03
Navigator	46	.63	5.82	5.10	1.82	.24

c = Completed Link trainer check.
i = Did not complete Link trainer check.

number of training requirements completed at the end of 6 weeks of operational training. Among the copilots of these crews, 803 were found to have taken the Aircrew Classification Tests. The correlation between the augmented pilot stanine of these copilots and the number of training requirements completed by their crews at the end of 6 weeks was 0.03. This was obtained by converting separate correlations by classes to equivalent z values and obtaining the weighted average. Separate correlations by classes ranged from 0.34 to —0.27.

In the Fourth Air Force, data were obtained at the Lead Crew School at AAF Muroc on the relation between B–29 crew bombing scores and the stanines of various crew members. The correlations between the bombardier, navigator and pilot stanines of the 89 copilots and crew average circular error were —0.11, —0.07 and —0.03 respectively. Where extreme bombing errors were excluded to give a normal aiming error score, the coefficients were —0.02, —0.06 and 0.00. The correlations between the stanines and percent of bombs on which gross errors were made were —0.09, 0.02 and —0.01. These coefficients show there was essentially a zero relationship between copilot stanines and bombing scores.

SUMMARY

Because the copilot served on bombardment crews merely as a sort of assistant to the airplane commander, it proved to be difficult to obtain adequate measures of his proficiency in combat type operations. Training personnel frequently did not know the copilots well enough to rate their performance. Thus it was not surprising that the estimates of proficiency obtained were less reliable and less useful in studies of the validity of aptitude scores or stanines, than similar criteria obtained for airplane commanders.

In general the stanines of copilots were found to bear no relation to their success in operational training, at least as measured by most of the criteria studied. Performance on the Link Trainer and certain ground school grades were exceptions. The navigator and bombardier stanines were found to predict academic success better than did the pilot stanine. While there was some evidence the pilot stanines of copilots predicted their Link Trainer Performance, the available data indicated such performance was better predicted by their navigator stanine.

CHAPTER SIX

Navigator

ANALYSIS OF DUTIES

Job Description

Information concerning the duties of the navigator in operational training was gathered by psychological research personnel in the Second, Third and Fourth Air Forces as a part of their mission to discover and evaluate criteria of navigator proficiency. The following brief description of the duties of the navigator in operational training is derived from observations, interviews with instructors, material found in Army Air Forces Training Standard 20-2-1, Army Air Forces Manual 35-1 and the training directives of the three Continental Air Forces concerned with the training of navigators.

The essential job of the navigator was to obtain and use such navigational aids as would help him bring the plane to its destination; to maintain throughout the flight a current record of where the airplane was and to furnish a log of the entire mission upon its completion. In the accomplishment of this assignment, the navigator was required to perform any one or all of the following specific tasks:

1. Make a preflight check of equipment.
2. Obtain thorough briefing instructions as to destination, course, time of arrival, expected areas of flak and hostile interception, and weather.
3. Navigate over land or sea by dead reckoning, pilotage, celestial means or radio.
4. Navigate by use of instruments such as driftmeter, pelorus, aircraft octant, radio compass, aperiodic compass and loran equipment.
5. Compute effects of various factors on course and plot projected course on the chart.
6. Maintain log.
7. Request radio bearings and make periodic position checks by this means.
8. Furnish other crew members data on wind direction and velocity, ground speed and drift on any heading.
9. Check data for errors after flight and recalibrate all navigation instruments if necessary.
10. Perform duties of bombardier, gunner, copilot or radio operator in emergency.
11. Send and receive radio, telegraph and blinker code.

In addition to these particular duties the navigator in operational training was required to provide the airplane commander and crew with technical information about machine guns, deflection problems, and the effect of weather on military operations. In some crews, the navigator was designated as crew personnel equipment officer.

Job Specifications

In determining the skills, knowledge, etc., considered desirable for the navigator, information was obtained from the following sources: Flying Evaluation Board case records, interviews, and questionnaires.

Flying Evaluation Board Records

For the period from January to October 1944, Fourth Air Force Flying Evaluation Board reports on 15 navigators revealed 3 major categories of causes for referral to local boards. Table 6.1 lists these categories and the frequencies with which they were cited as principal cause for Flying Evaluation Board action.

TABLE 6.1.—*Principal causes of flying evaluation board action on navigators*
JANUARY TO OCTOBER 1944—FOURTH AIR FORCE

	Number
Cause of action:	
Physical disability	2
Lack of proficiency	3
Personality defect:	
Fear and/or neurosis	8
Personal misconduct	1
Lack of maturity	1
Total	15

Of the 16 cases of navigators appearing before Flying Evaluation Boards in the Third Air Force during 1944, all but 1 were re-evaluated because of "personality difficulties." Since the term covered a wide range of causes, this result provides little insight, except to corroborate the Fourth Air Force finding that personality factors, rather than lack of proficiency were responsible for most of the failures in training. Due to the small number of cases and the inadequacy of records, Flying Evaluation Board proceedings were not particularly fruitful in providing information about the desirable characteristics of navigators.

Interviews

In an attempt to ascertain the characteristics of successful lead crew navigators, aviation psychologists interviewed returnee pilots, bombardiers and navigators who had flown in squadron, group and wing lead positions in combat. The consensus of these officers from different theaters as to the qualifications needed for each crew member was determined by analysis of these interviews. No attempt was made to scale the qualifications relative to their importance for lead crew position in combat. The following is a

list of the qualifications obtained in the Second Air Force by this interview technique:

1. General high proficiency in all phases of navigation.
2. Very high proficiency in celestial navigation, (Pacific Theaters, mainly).
3. Very high proficiency in pin-point navigation.
4. Knowledge of formation flying.
5. Foresight and anticipation in formation flying.

Questionnaires

On the basis of observations during training flights, examination of training directives and informal talks with flight instructors, a list of 10 factors considered important for the successful completion of Fourth Air Force training was obtained. Instructors were then asked to rank these items in order of importance and after completion of the ranking, to suggest additional factors which they considered important for the successful navigator. The reliability of instructors' judgments of the relative importance of items was determined by combining the rankings made by instructors at all three bases, correlating the average rank assigned by odd instructors with that assigned by even instructors (regarding items as individuals) and adjusting the coefficient by the Spearman-Brown formula. With 17 instructors, the reliability coefficient thus obtained was 0.88. In other words, the hierarchy as a whole was highly reliable.

TABLE 6.2.—*Order of importance of items as ranked by instructors and as indicated by correlation between item ratings and over-all ratings*

B-24 NAVIGATORS—FOURTH AIR FORCE

Item	Mean rank given by instructors	S. D.	Order of importance in terms of: Instructors ranking	Correlation between ratings
Ability in DR navigation	2.85	1.79	1	2
Eagerness	3.88	2.91	2	3.5
Ability in celestial navigation	4.00	2.17	3	1
Ability in pilotage	4.56	1.75	4	5
Knowledge of navigation equipment	4.94	2.42	5	6
Performance in emergencies	5.12	2.84	6	3.5
Ability to get along with crew	6.53	2.59	7	8
Ability in radio navigation	6.59	1.78	8	9.5
Conscientiousness of log procedure	7.06	1.66	9	9.5
Likeableness	9.47	1.65	10	10

In table 6.2 are given the average ranks given each item together with the standard deviation of the ranks. The latter are important in that they indicate the extent of agreement among instructors as to the importance or lack of importance of the various items. There was relatively good agreement that ability in D R navigation was most important, that ability in pilotage was about middle in importance and that conscientiousness of log procedure, ability in radio navigation and likeableness were of least importance. The lack of importance attributed to radio navigation may have been due either to a belief that the procedures were so simple navigators had little trouble with them, or to the fact that its use was contingent upon

the presence of radio installations, which were not then plentiful in the combat area for which these navigators were being trained. It is interesting that conscientiousness in log procedure, which was greatly stressed in training, was thought of such little importance. As with the airplane commander in chapter 4, "eagerness" or motivation was considered the most important *trait* or personal characteristic, although there was less agreement among instructors than for the other items mentioned above.

Navigation instructors were also asked to rate each navigator they had checked out as "above" or "below average" with respect to each of the factors described above. In addition, they rated each navigator as to over-all proficiency and ability. Separate sheets of paper each with a list of the navigators in the Section were used for each item in order to minimize halo effects. By comparing the importance of each item as expressed in the ranking of the items by instructors with the relative degree of correlation between ratings of navigators on the item and their over-all ratings, it was possible to see how much agreement there was between these two ways of estimating trait importance. In table 6.2 are shown the 10 items listed in order of their importance as ranked by the instructors. Also shown is the relative importance in terms of size of correlation between ratings on the items and over-all ratings. There is seen to be good agreement between what traits the instructors said were most important when they ranked the items and what traits were considered most important when they rated the navigators.

Additional items suggested by instructors as being important to training success were in the order of frequency of mention: general intelligence, maturity, ability to assume responsibility and cooperation with bombardier.

CRITERIA

Particular Criteria

Mission Grades

Navigators entering B–29 Combat Crew Training Station (CCTS) training in the Second Air Force were given a special 10-day course in over-water navigation at AAF Jackson, Miss. Usually this course was given prior to regular CCTS crew training. Each navigator was expected to fly three standard missions which were selected from a group of nine missions, six over water, and three over land, the latter being flown when weather did not permit over-water flights. Grading of missions was in terms of a 3-point scale, with points assigned as follows: Very Satisfactory, 3 points; Satisfactory, 2 points; and Unsatisfactory, 1 point. Reliabilities of navigation mission grades are shown in table 6.3.

In the Fourth Air Force, also, from time to time, in some sections at some bases, student navigators rendered reports on their navigation flights. The following kinds of errors were reported:

1. Average error of lines of position (LOP).

2. Average hourly error of estimated time of arrival (ETA).
3. Course error per 60 miles.

These reports were subject to an indeterminate amount of distortion because the reports were made out with the assistance of other crew members. From base to base within the Fourth Air Force, from section to section at the same base, and from month to month at the same base there was no consistent policy requiring the submission of mission reports. Mission reports, therefore, as used in the Fourth Air Force, were of practically no value as criteria.

TABLE 6.3.—*Reliabilities of mission grades and error scores*

B-29 NAVIGATORS—JACKSON AAF—SECOND AIR FORCE

Missions compared	Number of cases	Mean, *1*	S. D., *1*	Mean, *2*	S. D., *2*	r_{tt}
Mission grades:						
First v. second	443	1.93	0.66	2.05	0.68	0.16
First v. third	430	1.91	.66	2.13	.62	.20
Second v. third	430	2.08	.67	2.13	.62	.33
ETA error:						
First v. second	595	3.71	3.3	3.76	3.1	—.02
First v. third	555	3.85	2.9	3.73	2.9	.07
Second v. third	552	3.73	3.1	3.61	3.1	.13
Course error:						
First v second	616	9.50	6.7	8.35	6.3	.05
First v. third	536	9.57	6.7	8.11	6.1	.06
Second v. third	527	8.41	6.2	8.04	6.3	—.06

Mission Error Scores

In addition to the over-all grade assigned a navigator for the missions flown, records were maintained at AAF Jackson of the ETA (estimated time of arrival) and course errors for each leg of the mission. Number of legs varied from one to four depending upon the particular mission being flown. In general, position was corrected by the instructor at the start of each leg so that cumulative errors were not involved in these measures. In table 6.3 are given the correlations between average ETA and course errors of different missions. The ETA errors were recorded in minutes and the course errors in miles. It can be seen that the objective error scores of different missions had almost zero reliability. There are of course many sources of variability responsible for this outcome. Some of these are weather, differences in difficulty of missions, differences in emphases on different missions and instrumental errors. Space does not permit of an exhaustive analysis in this volume. A more detailed account of factors influencing these errors may be found in Report No. 10, Psychological Research on Navigation Training.

Log Grades

An examination of Second Air Force student navigator logs revealed that the usual grade given was "satisfactory." Occasionally, a log was marked unsatisfactory, incomplete, or not graded. It was believed, after discussions with staff navigators, that log grades were not a good representation of the

instructors' judgments of the quality of navigation on the missions. Several logs were examined by the Psychological Section, Hq., AAF Training Command, to determine whether any method of objective scoring could be devised. It was found possible to estimate fairly accurately the true flight path from examination of all the logs of a formation. It was then possible to assign more or less objective grades to the individual logs upon the basis of deviation from the true path. However, since the amount of labor involved in this procedure made this type of analysis almost impossible to use on a large scale, further studies of this type were abandoned. Considerable inference was also involved in the procedure concerning which planes actually constituted the formation. This information was not found in the logs.

In the Third Air Force the log of each navigator for every training flight was graded. Log grades were based upon completeness, neatness and exactness of computations. The care and objectivity of grading varied from station to station. For 23 cases at Barksdale Field, Class 0911, the odd-even log grade reliability was found to be 0.70.

The lack of uniformity of grading procedures makes it difficult to evaluate the potentialities of the log grade as a criterion. In its present unstandardized state this measure is of little value as an index of proficiency in navigation training in the Continental Air Forces.

Ratings

Two sets of Second Air Force navigator instructors ranked the student navigators in their flights in the order of their expected proficiency in combat. For one class of 56 student navigators the correlation between the 2 sets of rankings was 0.63. For another class of 56 student navigators, efficiency ratings (Form 66-2) for 2 periods of training were available. The correlation between these two sets of ratings was found to be 0.35. Additional evidence regarding the degree of reliability of subjective ratings of navigators can be obtained from an examination of the correlations between different sets of ratings as shown in tables 6.4, 6.5 and 6.6.

The efficiency and officer quality ratings shown in table 6.4 were ratings by instructors on a five-point scale: Superior, excellent, very satisfactory, satisfactory and unsatisfactory. Actually, observers rarely used more than the top three ratings. The ratings by instructors on the three qualities of AFTRC Form 2 were on a three point scale: above average, average and below average. Rank in class was based on a listing of officers in order of anticipated proficiency in combat. These ratings may or may not have been done by the same observer. The interview ratings in table 6.5 were on 8 qualities using a 10-point scale, and represent the consensus of the interview board. The sum of these ratings was used for interview board ratings in table 6.6. The flight surgeon ratings on the same table were on a three point scale: above average, average and below. Crew ratings were the average of ratings on a 10-point scale on 9 traits by three other members

of the crew. Self ratings were the sum of ratings on 9 qualities, using a 10-point scale for each. Thus all of the ratings in table 6.6 were made by different individuals and were relatively independent.

Among the factors entering into all ratings and rankings are the experience and prestige of the individuals being rated. One would, therefore, expect that first lieutenants and captains would be given higher ratings than would flight officers and second lieutenants. From the ratings in table 6.6,

TABLE 6.4.—*Correlations between different ratings of student navigators*

B-17 NAVIGATORS—SECOND AIR FORCE

Type of rating	Code	A	B	C	D	E	F	G
Efficiency ratings, Form 66-2	A		.....	.....	0.23	0.54	0.37	0.60
Officer quality ratings	B	.....		0.58	.80	.73	.62	.....
Technical aircrew skill	C	.....	55		.57	.54	.57	.....
Rating AFTRC Form 2:								
Military discipline	D	38	49	49		.65	.38	.06
Manner of performing duties	E	38	49	49	82		.44	.09
General aircrew ability	F	38	49	49	82	82		.42
Rank in class	G	167	.....	.....	38	38	38	

GENERAL NOTE—Entries below the diagonal are the number of cases for the corresponding correlation coefficients.

TABLE 6.5.—*Correlations between interview board ratings of seven traits*

B-24 NAVIGATORS—10-POINT SCALE—SECOND AIR FORCE

Trait	Code[1]	A	B	C	D	E	F	G
Enthusiasm for job	A	.....	0.42	0.41	0.29	0.55	0.35	0.36
Desire for combat	B	.....	.....	.36	.25	.31	.08	.06
Interest in flying	C	.....	.....	.....	.48	.58	.04	.36
Getting along with others	D	.....	.....	.....	.....	.62	.06	.08
Officer qualities	E	.....	.....	.....	.....	.....	.39	.36
Educational background	F	.....	.....	.....	.....	.....	.....	.30
Work experience	G	.....	.....	.....	.....	.....	.....	.....

[1] N = 54.

TABLE 6.6.—*Correlations between different ratings*

B-24 NAVIGATORS—SECOND AIR FORCE

Type of rating	Code[1]	A	B	C	D	E
Interview rating	A	.....	−0.41	−0.16	−0.05	0.36
Flight surgeon's rating	B	.....	.....	.15	.00	.00
Crew rating	C	.....	.....	.....	.22	.12
Self rating	D	.....	.....	.....	.....	.24
Rank in class	E	.....	.....	.....	.....	.....

[1] N = 54.

a composite rating score was prepared for each navigator and the average composite score determined for two groups: those first lieutenant and above in rank, and those second lieutenant and below. The biserial coefficient of correlation between composite ratings and rank was 0.45 from the total group of 52 navigators. However, not all of this correlation was necessarily due to "halo" since to a certain extent higher rank was a reflection of additional experience and ability.

In an attempt to develop an instrument for the evaluation of B-29 lead crew proficiency in the Second Air Force, two descriptive rating scales, Scale A and Scale B were developed. Both scales consisted of a number of

items upon which ratings from 1 to 5 were to be made, 1 in each case being the most favorable rating and 5 the poorest rating. Some of the items were the same for all crew positions. Others varied from position to position. Scale A was administered at Clovis, Alamogordo, Davis-Monthan and Biggs, and ratings with this scale were obtained from instructors for 160 navigators. Ninety-two navigators from AAF Albuquerque, Pyote, Walker and Great Bend were rated by their instructors on Scale B. In each case ratings were obtained from at least two independent observers for each of the individuals and crews rated. The correlations between independent ratings on both scales are shown in table 6.7.

TABLE 6.7—*Reliability coefficients of rating scale total scores*

SCALES A AND B—SECOND AIR FORCE

	N	First rater, 1		Second rater, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Scale A:						
Alamogordo 3	23	2.57	0.49	2.54	0.39	0.24
Alamogordo 6	23	2.16	.55	2.34	.49	.08
Biggs	41	2.86	.57	2.78	.57	.88
Clovis	36	2.97	.61	2.67	.62	.62
Tucson	37	2.92	.52	2.69	.71	.74
Combined (By Fisher's z)	160	2.76	.62	2.64	.61	.66
Scale B:						
Walker	49	2.51	.86	2.30	.82	.33
Albuquerque	19	2.67	1.01	2.40	.82	.58
Pyote	24	2.77	.88	2.53	.66	.60
Combined (By Fisher's z)	92	2.63	.90	2.38	.78	.46

At all bomber bases in the Fourth Air Force, performance ratings of student navigators were made by the squadron navigator in conference with flight instructors. Although the ratings were made upon the conventional army 5-point scale, rarely were more than 3 points actually used, and in most cases approximately 90 percent of all students were placed at the "Excellent" point on the scale. The lack of spread in these ratings made them practically useless for statistical analysis.

A rating scale consisting of 11 items was administered to Fourth Air Force B-24 navigator instructors at Muroc, Walla Walla and March Field. (A later and greatly modified B-24 scale was developed too late for administration.) The items for ratings were selected upon the basis of observations during training flights, examination of training directives, and informal talks with flight instructors. In response to the items, instructors were requested to rate (half above-average, half below-average) all the students whom they had check-ridden. In order to minimize "halo" effect, the items were administered successively on separate sheets of paper upon which were typed the names of all students in the section. Item reliabilities in the rating situation could be compited for only one sample of 21 Muroc student navigators for whom two or more ratings were available. The items of the rating scale, together with their reliabilities (r_{tt}), appear in table 6.8.

Research personnel in the Fourth Air Force assisted in the construction of a rating scale for use in the Lead Crew School at Muroc Army Air Base. This scale required instructors to rate navigators on seven items: preparation for mission, technique, accuracy, scope interpretation, turn procedure, interphone procedure and bombing team coordination. In addition, the instructor was asked to make an over-all estimate of the navigator, particularly in terms of his potentialities as a lead navigator. Two scores were evaluated for reliability: the average grade (the average of the grades on the seven items) and the over-all grade. Table 6.9 shows the reliability coefficients that were obtained by correlating the average rating of the odd instructors with those of the even instructors, corrected by the Spearman-Brown formula. The considerable drop in reliability for both scores after the first class was thought to be in large part due to increasing administrative difficulties which curtailed greatly the number of ratings on each navigator.

TABLE 6.8.—*Reliabilities of navigator rating-scale items*

21 NAVIGATORS MUROC AAF—FOURTH AIR FORCE

Item:	Reliability (r_{oe})
1. What is your over-all rating of the student as a navigator?	0.40
2. How eager is the student to learn his job as navigator?	.40
3. How well does the student get along with his other crew members?	.22
4. How well does the student perform in an emergency situation?	.70
5. How likeable is the student as a person?	.00
6. How good is the student's knowledge of navigation equipment?	.00
7. How good is the student's DR navigation?	.09
8. How good is the student's celestial navigation?	.60
9. How good is the student's pilotage?	.20
10. How conscientious is the student's log procedure?	.75
11. How good is the student's radio navigation?	.15

TABLE 6.9—*Reliabilities of navigator ratings*

MUROC AAF—FOURTH AIR FORCE

Measure	June		July		August		Total	
	N	r_{oe}	N	r_{oe}	N	r_{oe}	N	r_{oe}
Average rating	23	0.91	38	0.48	37	0.02	98	0.41
Overall rating	23	.84	38	.18	37	.00	98	.27

Ground Trainer Grades

1. *Celestial navigator trainer.*—The Celestial Navigator Trainer (CNT) simulates many of the flying conditions which confront a navigator in actual flight. The trainer was so constructed as to provide objective measures of the students' performance from the graphic records of the course "flown" by the student. These measures were: number of celestial fixes, average celestial fix error, average error of estimated time of arrival (ETA), average

track error, number of radio fixes and average radio fix error. In practice, these objective scores were not always systematically recorded. Frequently each of the student's four to eight missions on the trainer was rated by an instructor who professed to take into account both the quality of the performance and the objective scores. In both the Second and Fourth Air Forces the missions were divided into the two categories of simple and complex.

The data in table 6.10 indicate the reliabilities of the ratings or grades and the reliabilities of objective scores were similar in magnitude, although the comparison is somewhat hazardous in view of the differences in numbers of cases, etc. Somewhat lower reliabilities for objective scores are reported in table 6.11 for 56 student navigators at Sioux City.

In this same study it was found that if only successful missions were included the reliability of the scores was not significantly affected.

TABLE 6.10.—*Reliability coefficients of CNT scores and ratings*

B-17 AND B-24 NAVIGATORS—SECOND AIR FORCE

Type of CNT grade or score	Number of cases	r_{tt}
First mission v. second mission grades	746	0.39
Average grade, odd v. even missions	812	.41
Average grade, first half v. second half training	142	[1].66
Average ETA error, odd v. even missions	36	.66
Average course error, odd v. even missions	36	.38
Average number of fixes, odd v. even missions	36	.71

[1] Not corrected by Spearman-Brown Formula.

TABLE 6.11.—*Reliability coefficients for CNT objective scores*

ODD-EVEN MISSIONS—SIOUX CITY—SECOND AIR FORCE

Score	N	Odd missions, 1		Even missions, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Number of fixes	56	3.89	0.60	3.98	0.74	0.56
Av. position error	56	8.31	2.14	8.53	2.26	.18
ETA error	56	2.48	1.55	2.83	1.52	.29
Distance error	56	7.70	4.18	6.62	3.62	.15

Table 6.12 presents reliability coefficients for CNT ratings from two other air forces. In general, these coefficients resemble those shown in table 6.10.

TABLE 6.12.—*Odd-even reliabilities of CNT ratings of navigators*

THIRD AND FOURTH AIR FORCES

Base	Air force	Number of cases	r_{tt}
Walla Walla	4th	56	0.53
Muroc	4th	44	.11
March Field	4th	50	.28
Gulfport (Class 0731)	3rd	35	.64
Gulfport (Class 083[illegible])	3rd	45	.37

In the Fourth Air Force, correlations were obtained between objective scores and ratings, and the resulting coefficients are presented in table 6.13.

The results suggest that the two variables were as highly correlated as could be expected in view of the low reliability of each.

TABLE 6.13.—*Correlation between CNT objective scores and instructors ratings of CNT performance*

50 NAVIGATORS—FOURTH AIR FORCE

CNT Scores	*Correlation between objective scores and instructors' ratings*
Simple missons (61 missions):	
Number of celestial fixes	0.37
Average celestial fix error	.67
Number CF plus CF error (equally weighted)	.60
Complex missions (39 missons):	
Number of celestial fixes	.24
Average celestial fix error	.86
ETA error	.36
Track error	.52
Number CF, CF error, ETA error, track error, (equally weighted)	.76

The wide range of reliability coefficients for both types of scores, together with small samples for several types makes generalizations with regard to the adequacy of CNT scores difficult. However, since some degree of reliability was found, the various scores were regarded as of sufficient value to be useful for validation purposes.

2. *G-1 trainer.*—The G-1 Trainer provided training in dead reckoning. Student navigators in the Second Air Force were rated on a 6-point scale (0, very poor, to 5, excellent) for each of 5 to 10 missions. For a group of 560 student navigators an odd-even mission reliability of .28 was obtained for these ratings.

Academic Grades

Tests and grades covering various aspects of academic courses were in use at practically all training bases. There was, however, no uniformity from base to base either in the construction of the tests and their subsequent use or in grading procedure. Many of the tests were administered but never graded. Some were graded but the grades were not recorded. In many cases the tests served "to encourage study" or "to help the student review." In the majority of academic courses taken during navigator training the "grade" consisted of a notation that the student had attended the required number of hours for completing the course.

A split-half reliability of 0.18 for navigation grades was obtained for 56 cases in the Second Air Force. For a similar group in a bombardment course the correlation between first-half and second-half grades was —0.12. The average intercorrelation of three phases of gunnery grades was 0.60. The tests upon which these grades were based were not available for examination. Lack of uniformity in construction, administration and grad-

ing of tests in academic subjects makes it difficult to evaluate academic grades.

Crew ACE

A detailed description of this criterion is given in Chapter 7 in the section on bombing scores. The use of bombing error scores as criteria of navigator proficiency requires a special word, however. It has been stated repeatedly that bombing scores reflect to a certain extent the ability of all members of the crew. The navigator was, however, a minor assistant in the visual bombing task. Radar bombing permitted the navigator to play a more important role. The navigator was equipped with a radar scope and in some crews he actually performed the duties of the radar observer and/or bombardier during the bombing procedure. There was, however, no uniformity from crew to crew with respect to the division of duties among the members of the radar bombing team, hence there was no way of determining the extent to which the navigator's proficiency was reflected in crew radar ACE.

Because of the possibility that a navigator's proficiency might affect the accuracy of a crew's bombing, the radar ACE score was regarded as a potentially useful criterion, especially since it was moderately reliable. For example, radar bombing data obtained at Muroc Army Air Field had an odd-even mission reliability coefficient of .69 for 102 cases. In the Second Air Force the odd-even mission reliability coefficient was 0.61 for 219 cases.

Flying Evaluation Board Records

As stated above, records of Flying Evaluation Board proceedings were not regarded as suitable for criterion purposes because of the small number of cases and the inadequacy of the data recorded.

Accomplishment of Training Requirements

In the Second Air Force, it was possible to obtain the mission and training accomplishment scores for 1,200 crews. The scores were the number of training requirements completed at the end of 6 weeks. This score was subject to a number of criticisms as a criterion of navigator proficiency. The rate at which a crew accomplished its training responsibilities was affected by many factors. The ability of the navigator was probably not a major determinant. It was not possible to determine the reliability of the scores.

Intercorrelations

Tables 6.14, 6.15, 6.16 and 6.17 present intercorrelations among measures of navigator proficiency in each air force in which navigators were trained. The results are not comparable from table to table since different scores are compared and in the case of similar scores different conditions doubtless obtained. However, the general impression is one of little or no correlation between the scores. The results presented in table 6.16 appear

to contradict this generalization, since the bombing scores were moderately intercorrelated as were the rating scores (to some extent). It should be noted, however, that the bombing scores were actually different methods of evaluating the same bombing performance. The correlation between rating scale scores is doubtless due to "halo," since the different scores were prob-

TABLE 6.14.—*Correlations between different types of navigation criteria*

B-17, B-24 AND B-29 CREWS—SECOND AIR FORCE

Types of criteria compared	Number of cases	r
Rank in class v. academic average	150	0.14
Rank in class v. navigation proficiency test	56	.22
Rank in class v. CNT grade	56	.08
Efficiency rating, form 66-2 v. academic average	38	.19
CNT grade v. navigation proficiency test	56	.02
CNT grade v. dead reckoning trainer grades	525	.20
Rating scale scores (scale B) v. radar CE T-scores	86	−.04
Rating scale scores (scale A) v. ground school grades	160	.22
Rating scale scores (scale A) v. CNT grades	160	.08
Radar ACE, T-scores v. ground school grades	125	−.22

TABLE 6.15.—*Intercorrelations among measures of navigator proficiency*

BARKSDALE FIELD—THIRD AIR FORCE

	CNT	Ground school	Log grades	Class ranking
CNT		0.21	0.30	0.40
Ground school	43		.02	.15
Log grades	40	39		.00
Class ranking	41	40	40	

GENERAL NOTE—Entries below the diagonal are the numbers of cases for the corresponding correlation coefficients.

TABLE 6.16.—*Intercorrelations among measures of navigator proficiency*

B-24 NAVIGATORS—FOURTH AIR FORCE

	Code	A	B	C	D	E	F	G
1. Over-all rating	A		0.26	0.10	0.28	0.09	−0.21	−0.33
2. CNT rating	B	145		−.11	−.08	.06	−.15	.00
3. Ground school grades	C	145	147		−.01	.12	−.30	−.37
4. Log grade	D	54	56	54		......	......	......
5. G-1 trainer rating	E	40	40	40			−.25	.01
6. Average circular error	F	40	40	40	......	40		.47
7. A-2 circular error	G	40	40	40	......	40	40	

GENERAL NOTE—Entries below the diagonal are the number of cases for the corresponding correlation coefficients.

TABLE 6.17.—*Correlations among measures of navigator proficiency*

B-29 CREWS—FOURTH AIR FORCE

	Code[1]	A	B	C	D	E
ACE total	A	......	0.74	0.87	0.19	0.30
ACE normal aim	B	......	......	.45	.17	.23
Per cent gross errors	C	......	......	......	.19	.26
Average rating	D	......	......	......	......	.78
Over-all rating	E	......	......	......	......	......

[1] N = 102.

ably assigned by the same instructors. The correlations between rating scores and bombing scores are low and corroborate the low intercorrelations found in other tables. The low intercorrelations among the measures of navigator proficiency are doubtless due in part to the low reliabilities of most of the variables.

Evaluation of Criteria

Because at many bases mission grades were made out with the assistance of other crew members, these reports were subject to an indeterminate amount of distortion. From base to base, from section to section at the same base, and from month to month at the same and different bases there was no consistent policy requiring the submission of mission reports. For these reasons mission reports were of little value as criteria of success in navigator training. Even the more standardized grading of B–29 missions at Jackson AAF yield quite low reliability coefficients.

Lack of uniformity among and within bases in the maintenance and grading of logs makes it difficult to evaluate the potentialities of log grades. Although a method of estimating true flight path from an examination of all of the logs of a formation was developed and it was possible to assign relatively objective grades to the individual logs on the basis of indicated deviations from the true path, the amount of labor involved in this procedure makes this type of analysis almost impossible to use on a large scale. The fact that a moderately high reliability coefficient was obtained for a small sample in the Third Air Force suggests that log grades are potentially a valuable criterion, if grading procedures are standardized. Rating scales showed some promise as criteria of proficiency in navigator training. In general, whenever it was possible for psychological personnel to control, at least in part, the conditions under which scales of their own design were administered, and when ratings were given at the same base by individuals in the same administrative relationship to the men being rated, the rating scales yielded moderately high reliability coefficients. However, the influence of subjective factors makes them less satisfactory than would be indicated by the reliability coefficients obtained.

The value of CNT ratings and grades was reduced by the lack of reliability of both the objective and subjective measures. More careful recording of the "objective" measures and the insistence upon a standardized procedure for evaluating these measures might produce a grade or rating that would prove to be a satisfactory criterion of navigator training proficiency.

The present lack of uniformity in construction, administration and grading of tests of academic subjects made academic grades of little value as validation criteria. If pass-fail requirements could be enforced and if testing conditions could be standardized, there is little doubt that reliable tests of proficiency in academic courses could be devised.

The little work that has been done in determining the relationship of navigator's ability as measured by a rating scale and radar average circular error is promising. Such measures of proficiency will doubtless become more important as emphasis is placed on the performance of the crew as a whole.

The small number of Flying Evaluation Board cases and the inadequacy of the reports themselves as to possible reasons for reevaluation render this source of little value as a criterion.

Two general observations might be made in connection with evaluation of all the criteria. First, it is obvious that for practically all the available criteria it was difficult to obtain comparable data from station to station and from time to time at the same station. Secondly, it must be noted that there was no simple administrative procedure for eliminating students who lacked proficiency in their navigator training. Because of this lack, training personnel saw little value in trying to make fine distinctions as to proficiency. This, in turn, made it difficult to introduce new measures of proficiency or to make much progress in refining existing measures.

VALIDATION STUDIES

Mission Error Scores

Unfortunately, no validity data were available for mission grades at AAF Jackson in the Second Air Force. However, the ETA and Course Error scores obtained were compared with the navigator stanines of the navigators concerned. As would be expected from the nearly zero reliability of these scores, the validity coefficients of the stanines against these criteria were not significantly different from zero.

Log Grades

Data from both the Second and Third Air Forces indicate little relationship between the stanines of navigators and navigation log grades. For a sample of 53 cases at Biggs Field, the correlation between the navigator stanine and log grades was 0.10. For 44 B–29 navigators in class 0807 Barksdale Field, coefficients were —0.08, 0.03, 0.01 and —0.09 for the bombardier, navigator, pilot and pilot plus credit stanines, respectively.

Ratings

Table 6.18 lists the correlations between each of a number of ratings of student navigators in the Second Air Force and navigator stanines. The efficiency ratings, ratings of attitude toward work, officer quality and technical aircrew skill were all five-point ratings of the usual type already discussed in Chapter IV and ranged from "unsatisfactory" to "superior." In practice, few ratings other than "very satisfactory" or "excellent" were given. Ratings of general aircrew ability and flight surgeon's rating were

on a three-point scale of above average, average and below average. Interview ratings, self ratings and crew ratings were 10-point ratings on a total of 8 traits or characteristics. Rankings were obtained by having instructors rank the navigators in their flights or classes in order of their expected proficiency in combat.

Table 6.18.—*Correlations between navigator stanine and ratings*

B-17 AND B-24 NAVIGATORS—SECOND AIR FORCE

Criterion	Number of cases	Navigator stanine, 1		Criterion, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Efficiency rating 66-2	156	7.37	1.15	41.1	11.5	[illegible]
Rankings by instructors	211	7.38	1.18	50.1	18.8	[illegible]
Technical aircrew skill	51	7.59	.99	42.9	9.6	[illegible]
Officer quality rating	51	7.59	.99	37.3	10.8	[illegible]
General aircrew ability	49	7.59	1.21	47.9	10.6	[illegible]
Attitude toward work	192	7.65	1.10	2.19	.4	**.[illegible]
Interview ratings	53	7.68	.90	5.99	3.2	−.11
Flight surgeon's rating	53	7.68	.90	2.02	.4	−.[illegible]
Self rating	53	7.68	.90	2.77	.9	.06
Crew rating	53	7.68	.90	7.02	2.4	−.[illegible]

None of the correlations between stanines and ratings of different types is significant except that for the "Attitude Toward Work" rating which yielded a correlation with navigator stanine significant at the 1 percent level. In addition to the rankings listed in table 6.18, rankings of Second Air Force student navigators by their instructors were secured from a number of CCTS classes. For a group of 721 student navigators these rankings correlated 0.04 with navigator stanine. Results presented in table 6.19 indicate that there was little or no relationship between the stanines of navigators and descriptive rating scale scores. In the case of Scale B, there is a low but statistically significant correlation between ratings and the pilot stanine of the navigator.

Table 6.19.—*Correlations between total rating scale scores and stanines of navigators*

SUM OF TWO RATERS, 16th AND 17th WINGS—SECOND AIR FORCE

Station	Stanine	N	Stanine, 1		Rating scale scores, 2		r_{12}	r_{12}[1]
			Mean	S. D.	Mean	S. D.		
Scale A:								
Alamogordo 3	N	22	7.45	0.84	20.68	2.82	0.38	
Alamogordo 6	N	20	7.50	1.07	18.00	3.02	−.02	
Biggs Field	N	37	7.51	1.00	23.32	4.10	−.06	
Clovis	N	35	7.54	.99	22.62	4.24	.00	
Tucson	N	36	7.42	.98	22.44	4.66	−.12	
All stations	N	150	7.48	.98	21.82	4.38	.00	0.01
All stations	B	150	6.75	1.55	21.82	4.38	.09	.06
All stations	P	150	5.97	1.75	21.82	4.38	.06	.04
Scale B:								
Pyote	N	46	7.52	1.12	28.98	7.48	.13	
Albuquerque	N	37	7.19	1.11	27.18	7.52	−.07	
Walker	N	42	7.69	1.19	25.22	6.48	.10	
Great Bend	N	37	7.81	1.09	25.46	8.96	.25	
All stations	N	162	7.56	1.15	26.78	7.20	.12	.11
All stations	B	162	6.71	1.62	26.78	7.20	.13	.14
All stations	P	162	5.91	1.67	26.78	7.20	**.21	**.21

[1] Coefficients were combined for the stations by converting into "z" equivalents and also by combining all data directly.

Ground-Trainer Grades

Correlations between navigator stanines and ground-trainer grades in the Second Air Force are presented in table 6.20. The navigator stanine is shown to have a significant relationship to CNT average mission grades. In the Third Air Force, on a sample of 60 B–29 navigators, no significant correlation was found between the CNT grades and the stanines; the coefficients being 0.18 for the bombardier, 0.03 for the navigator and 0.02 for the pilot stanine.

TABLE 6.20.—*Correlations between navigator stanines and ground trainer grades*

SECOND AIR FORCE

Criterion	Number of cases	Navigator stanine, 1		Trainer grade, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
CNT grade	54	7.31	1.41	78.2	5.4	0.14
CNT average mission grade	108	7.54	1.12	2.96	.5	**.30
Average DRT mission grade	53	7.68	.90	3.28	.4	.08

Academic Grades

Correlations between measures of academic success in the Second Air Force and navigator stanines are listed in table 6.21.

The navigator stanines appear to be significantly related to grades in the gunnery navigation examinations. Several other of the coefficients approach significance. In another study in the Second Air Force when 150 navigators were divided into "upper" and "lower" groups based on the total of all

TABLE 6.21.—*Correlations between navigator stanine and academic grades*

SECOND AIR FORCE

Criterion	Number of cases	Navigator stanine, 1		Academic grades, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Academic average	192	7.65	1.10	88.5	3.7	0.16
Ground school grades	105	7.39	1.20	46.9	11.0	.08
Gunnery grade	56	7.35	1.42	83.3	4.0	*.25
Bombardment grade	55	7.35	1.41	83.1	6.4	.11
Communications grade	55	7.34	1.41	97.4	1.2	−.03
Navigation grade	54	7.31	1.41	72.6	9.1	.02
Navigation examination	55	7.43	1.21	63.2	9.1	*.26

ground school grades, the navigator stanine showed no relationship to ground school performance. At Pyote Army Air Field for a class of 46 B–29 navigators the correlation between the navigator stanine and the navigation final examination was 0.47 which is significant at the 1 percent level. In the Third Air Force, for a class of 23 navigators, the navigator stanine correlated *0.45, significant at the 5 percent level, with average ground school grades.

Accomplishment of Training

There were 921 student navigators out of 1,200 heavy bombardment crews for whom mission and training accomplishment records as well as

stanines were available. The correlation between the navigator stanines and the number of training requirements completed at the end of 6 weeks by their crews was not significantly different from zero. The z coefficient for 24 classes with a total N of 921 was —0.03.

Crew ACR

The results of two studies comparing the navigator's stanine for his specialty with the radar "bombing" record of the crew of which he was a member are shown in table 6.22. The bombing scores of the three types occurring in the Second Air Force study were converted into T-scores for each station. These scores were then combined directly to obtain the coefficient for all the stations. In the data from the Fourth Air Force the separate coefficients for the three classes were combined in a weighted average using Fisher's z.

TABLE 6.22.—*Correlations between navigator stanine and crew radar ACR*

B-29 NAVIGATORS—SECOND AND FOURTH AIR FORCES

Station	N	Stanine, 1		ACR, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Great Bend	25	7.76	1.14	55.9	5.50	−0.26
Walker	29	7.79	1.22	50.4	3.98	.12
Alamogordo	13	7.28	1.64	49.6	4.44	−.39
All 2AF stations[1]	[illegible]	7.65	1.19	50.6	4.78	−.18
Muroc AAF (4AF):						
June	20	8.15	1.06	43.35	8.73	0.2
July	33	7.97	1.09	50.18	10.35	.39
August	38	7.22	.99	49.29	9.31	.12
Total Muroc[1]	91	......	......	......	......	.11

[1] Average by Fisher's z.

Radar bombing scores in the Fourth Air Force were analyzed in greater detail to yield two types of scores. "Gross errors" were extreme errors thought to be caused by failure to identify the target. Since the magnitude of such errors did not seem significant, the measure used was the percentage occurence of such errors (over 6,000 feet). The other score was the "normal aiming error" utilizing all scores with errors of less than 6,000 feet. With 91 navigators at AAF Muroc the validity of the navigator stanine in predicting these scores was 0.12 and 0.03 respectively, for the gross error and normal aiming error scores. Although not statistically significant, the larger coefficient obtained with the gross error score indicates this score may merit further study. One would expect that navigation errors might result in failure in finding the target.

In general, there appeared to be little evidence of relationship between navigator stanines and bombing measures.

Flying Evaluation Board Records

In both the Second and Third Air Forces, the stanines of navigators who met Flying Evaluation Boards were compared with non-FEB cases. In the Third Air Force, a sample of 13 eliminees was compared with a control

group of 13. No statistically significant differences were found. The same result was also found in the Second Air Force, in a larger study, comparing the stanines of 113 FEB navigators with a group of non-FEB navigators matching the former as to time of testing and training. The average stanine of eliminated navigators was 6.76 as compared with 6.96 for the controls, a difference not significantly different from what might be expected by chance.

In general, the stanines of navigators showed little relationship to the measures of navigator flight proficiency that were available. However, the navigator stanine was significantly related to several measures of ground school proficiency and to some ground trainer scores. These results call for some general comments concerning the circumstances under which the above research was conducted.

SUMMARY

Basic to the job of predicting success in navigational training in the Continental Air Forces is the problem of obtaining reliable and representative measures of the success that is to be predicted. Continental Air Force training was designed to meet military commitments rather than to provide a laboratory for psychological studies. Each station complied with basic training directives in a way which made it possible to meet its own requirements. From the point of view of training, variation among bases in compliance with training directives was justified on the basis of expediency. This variation was especially great in the recording of data concerning grades, ratings, etc., of almost every aspect of training. The difficulty of obtaining comparable data from station to station and from time to time at the same station made it very hard to obtain satisfactory criteria for validation purposes.

The problem of obtaining criteria was further complicated because no simple administrative procedures existed for the elimination of nonproficient students. The lack of such procedures made it difficult to convince training personnel that there was any value in measuring proficiency.

In the few situations where psychologists were able to adapt standard psychological procedures to the training situation and to administer such procedures under even moderately standardized conditions, it was possible to obtain reasonably reliable measures of proficiency. This would suggest that satisfactory criteria could probably be obtained if the pressure of training were relaxed sufficiently to permit the introduction of the controls and standardization lacking during the war. It is quite likely, however, that many apparently objective criteria of proficiency such as errors in estimated time of arrival, course errors and the like are influenced by so many variables that they cannot be relied upon, even under more standardized procedures. Unless reliable measures are developed, it is doubtful if the predictive efficiency of the present composite aptitude scores (stanines) for operational training can be materially raised.

One further comment concerning the validity coefficients obtained should be made. The range of stanines was severely restricted by the selection procedures in effect in the Training Command. The only trainees permitted to begin navigator training had extremely high stanines. Thus in operational training conditions were favorable for high stanine navigators to get low as well as high criterion scores and thus attenuate the validity coefficients.

CHAPTER SEVEN

Bombardier

ANALYSIS OF DUTIES

Job Description

To supplement the formal job analysis of the bombardier made by Psychological Research Project (Bombardier) (found in Report Number 9: Psychological Research on Bombardier Training) aviation psychologists in the Continental Air Forces gathered information on the job of the bombardier in operational training from a number of sources. Interviews with instructors and bombardiers with combat experience, AAF Training Standard 20-2-1 and directives of individual air forces provided a broad orientation to the nature of the task which was sufficient to meet the needs of the research sections. The following is a brief digest of some of the material from these sources.

The ultimate task of the bombardier was to locate, identify and accurately bomb assigned targets from a bombardment airplane. In order to accomplish this the bombardier had to adjust his bombsight and related equipment according to specific conditions such as ground speed, altitude and drift of the airplane. In addition, the bombardier was required to fire flexible guns and cannon and to make minor repairs on this equipment. The bombardier was also required to navigate by dead reckoning with a maximum course error of 1½° and a maximum ETA error of 1½ minutes per hour of flight.

Operational training programs in bombing, while they varied slightly in detail among the air forces, stressed training flights which established the bombardier's proficiency in solving the bombing problems under all conditions encountered in combat operations, including daylight and darkness, high and medium altitudes, singly or in formation, and long and short approaches. Training schedules required the bombardier to drop demolition and incendiary bombs, and to make a certain number of releases using the pilot direction indicator.

Job Specifications

The qualifications essential to success as a bombardier were investigated in a number of ways, the most fruitful of which were interviews with training personnel and analyses of existing rating methods. For the purpose at hand, Flying Evaluation Board reports (q.v. under CRITERIA) proved relatively worthless. Although lists of desirable characteristics varied some-

what by air force and by type of aircraft, many of the discrepancies either merely reflected differences in emphasis or were purely a matter of definition.

In the Fourth Air Force, job specifications of B–24 bombardiers were investigated principally through the use of a questionnaire. Nine items describing particular skills and personality traits were developed on the basis of observations during training flights, examination of training directives, and informal interviews with flight instructors. A questionnaire composed of the nine items was administered individually to nine flight instructors at each of three stations. Instructors were asked to rank the items in the order of their importance to successful completion of Fourth Air Force training. The reliability of the hierarchy as a whole was estimated by correlating average rank assigned by odd-instructors with that assigned by even-instructors, regarding the items as individuals. Corrected for total number of ratings by the Spearman-Brown formula, the reliability coefficient thus obtained was .61.

In table 7.1 the items are listed in the order of the average rank given them by the instructors. The average rank and standard deviation of ranks given are shown for each item.

TABLE 7.1.—*Relative importance of bombardier traits and skills*

BASED ON TRAIT RANKINGS AND RATINGS BY FOURTH AIR FORCE INSTRUCTORS

Trait or skill	Mean rank given by instructors	S. D.	Order in terms of rankings	Order of importance in terms of correlation with over-all rating
Good bombsight procedure	3.11	1.60	1	1
Ability to identify target	3.44	2.12	2	4
Eagerness	3.89	2.09	3	3
Knowledge of computers	4.00	2.11	4	3
Performance in emergencies	4.22	2.53	5	5
Ability to get along with crew	5.50	2.45	6	6.5
Ability to fly on C-1 autopilot	6.33	1.58	7	8
Likeableness	7.17	2.27	8	6.5
Good evasive action	7.33	1.78	9	9

The standard deviations of the rankings provide a convenient means of expressing the degree of agreement among the instructors as to the relative rank of an item. From the data in table 7.1 it is seen that instructors agreed well among themselves as to the importance or lack of importance of good bombsight procedure and ability to fly on C-1 autopilot. On the other hand, there was much less agreement as to the importance of performance in emergencies and ability to get along with the crew. Agreement on the other items was somewhere between these two extremes. If the items in table 7.1 are divided into skills and traits or attitudes, the relative importance of the items becomes more meaningful. Certain technical skills, such as good bombsight procedure, ability to identify targets, and familiarity with computers, were believed of paramount importance to the bombardier. Certain other technical skills were relatively unimportant to training success. Ability to fly on C-1 autopilot was thought of little importance, presumably because accurate bombing was possible, although

probably difficult, without its use. Evasive action was disparaged as a matter of training policy, since combat requirements in evasive action varied widely.

So far as traits and attitudes are concerned, motivation ("eagerness") alone seems to have been considered of essential importance. Adequacy during emergencies was somewhat less important than for airplane commanders (*see* chapter 4), probably because bombardiers were less frequently involved in major emergencies. Ability to get along with other crew members was thought no more significant for bombardiers than for airplane commanders, and likeableness was similarly regarded as relatively unimportant.

Ratings on these nine items, together with ratings of over-all proficiency, were also obtained from a number of instructors for bombardiers whom they had checked on one or more flights. By comparing the ratings given a bombardier on the different items with his over-all rating it was possible to see which of the nine traits or characteristics entered most into the over-all rating. This could be expressed statistically by the size of the coefficient of correlation between item and over-all rating. The larger the coefficient, the more important the item was in the over-all rating. In table 7.1, in addition to the order of rank of the items as ranked by the instructors, there is shown the relative rank of each item in order of the degree of correlation between ratings of bombardiers on the item and over-all ratings of the same bombardiers. It can be seen that there is good agreement between the instructors' beliefs as to the importance of the items (expressed in their rankings of items) and the unconscious importance given to the traits and characteristics in their ratings of bombardiers.

Qualifications regarded by instructors as desirable in B–29 bombardiers may be approximately determined from a rating scale used at the Lead Crew School, Muroc Army Air Field, Fourth Air Force. Table 7.2 shows the differential weights assigned to the 14 items in the scale. It will be noted that training personnel responsible for the scale entirely omitted personality items.

Job specifications for the B–17 bombardier in the Third Air Force were determined principally from interviews with training personnel. While tech-

TABLE 7.2.—*Items and weights for bombardier rating scale*

LEAD CREW SCHOOL, MUROC AAF, FOURTH AIR FORCE

Items weighted 1 (3 items)

Personal equipment.
Camera equipment.
Setting and preflighting altimeters.

Items weighted 3 (4 items)

Bombardier's kit and target folder.
Bomb racks, fuses, bomb bay doors.
Preflight of bombsight, autopilot, sighting station.
Ability as aerial observer.

Items weighted 10 (7 items)

Ability in using computers.
Preset data.
Identification of target.
Visual bombing procedure.
Radar bombing procedure.
Knowledge of [illegible]
Knowledge and execution of standard bombing procedure.

nical skills (e.g., smooth bombsight procedure and accurate target identification) were regarded as highly important, instructors assigned crucial significance to certain personality factors in discriminating "good" and "poor" bombardiers. These included: cooperative attitude, emotional stability, common sense or judgment, and motivation and initiative.

As with other aircrew specialties, study of the reports of Flying Evaluation Boards did not yield much information as to the real reasons for failure of bombardiers in operational training. Usually the reports permitted only a classification of reasons for re-evaluation into three major categories: personality defect (usually fear of flying), lack of proficiency and physical disability. The frequencies with which these categories appeared as reasons for Board action are given in table 7.11 under *Criteria*, a later section of this chapter.

CRITERIA

Average Circular Error

Practice Bombing

Average circular (radial) error for B-17 and B-24 crews was obtained from actual practice-bomb releases over targets marked on the ground, usually in the form of a cross with concentric rings for estimating distance of the hit from the center. As a routine matter circular error scores were approximately standardized for altitude by applying empirically determined corrections. Hence, the corrected scores approximately described the circular errors which would have resulted had the bombing been done at a common altitude (12,000 feet). The resultant scores were the product of many factors. Some of these factors reflected the true bombing proficiency of crews. Others, however, represented extraneous factors which were in varying degrees uncontrolled. The latter included differences in bombing altitude, differences in illumination and associated conditions (day *v.* night bombing), differences in scoring procedures (photographed *v.* estimated strikes), and differences in bombing procedure (C-1 autopilot *v.* manual releases). Differences between day and night bombing were actually not so crucial as might have been expected. For one sample of 51 cases in the Second Air Force the mean ACE for the day bombing was 256 feet with a standard deviation of 60 feet. For 49 cases of the same sample the mean ACE for night bombing was 275 feet with a standard deviation of 90 feet. Since there was no evidence that this factor seriously attenuated the reliability of average circular error, no attempt was made to correct circular error for differences attributable to this factor. Furthermore, for one class of 41 Fourth Air Force (Muroc) bombardiers, day *v.* night circular error correlated 0.43, a figure which compares favorably with uncorrected reliability coefficients for all circular error scores computed from odd-even missions.

Circular error was more seriously affected by the method of scoring

employed. Scores determined from strike photographs were usually larger than those based on bombardiers' estimates. Representative distribution statistics of average circular error for both methods of scoring are presented in table 7.3. Data from the Second Air Force tend to indicate that the inclusion of estimated strikes seriously reduced the reliability of average circular error of these crews was .86 as compared with a coefficient of 0.66 graphing all of their strikes. The odd-even mission reliability of the average circular error of these crews was .96 as compared with a coefficient of 0.66 for the total class. Furthermore, among the 6 classes (280 students) studied at another station (Ardmore), one class had photographed only 24 percent of the bomb strikes. The odd-even mission reliability of average circular error for this class was only 0.25, as compared with a coefficient of 0.56 for the total group of 6 classes in which approximately 50 percent of the strikes were photographed. However, the evidence on the amount of error

TABLE 7.3.—*Comparison of average circular error for photographed and estimated strikes*

Method	2d Air Force			4th Air Force		
	N	M	S. D.	N	M	S. D.
Photographed	52	220'	52'	51	343'	85'
Estimated	51	240'	75'	51	268'	70'

introduced by combining photographed and estimated strikes is not conclusive, since the samples studied were small and the influence of other factors was not controlled. In any case, no correction was applied for this factor in the research studies of average circular error.

The results obtained suggest that, although the bombardier gave himself the benefit of the doubt when he estimated his circular error, either his estimate was affected by what he knew his photographed performances were in the past, or else he did not completely ignore the objective facts of the performance in question.

Differences in bombing procedure constituted another source of variability in the average circular error. Bombing done with C-1 autopilot, with the plane being flown under the automatic controls was distinctly more accurate than manual bombing. In manual bombing, the plane was flown with the pilot manipulating the controls on the bombing run. The mean average circular error for a class of 51 Second Air Force bombardiers was 248 feet for releases under automatic control and 293 feet for releases under manual control with standard deviations of 57 feet and 62 feet respectively. Since from 60 to 95 percent of all bombing was automatic, no correction was employed to equate for differences in procedure. Failure to correct for these procedural differences was believed relatively unimportant from a practical standpoint in studies of reliability. Of course, for validation purposes corrected scores would have been preferable. However, it was not possible to obtain samples of sufficient size to permit the calculation of a correction factor in which any confidence could be placed.

The over-all reliability of average circular error for practice bombing can be estimated in a number of ways. Table 7.4 lists reliability coefficients obtained from various methods of pairing circular error data. It is believed that the most accurate estimate is obtained by correlating odd and even mission scores. Reliabilities computed from odd and even bomb releases are spuriously high since odd and even releases are really parts of the same performance. The variables affecting bombing scores, such as aircraft, motivation, weather, etc., are here virtually constant. Correlation of scores from first half and second half of training yielded spuriously low coefficients because of differences in type of mission, training directives, effects of learning and a host of other factors.

TABLE 7.4.—*Reliability coefficients for average circular error in practice bombing*

Method	Air force	Station	N	Number of missions	r_e
Odd v. even missions	Second	Ardmore	280		0.56
Do	Second	Biggs[1]	53		.51
Do	Second	Pueblo	56		.66
Do	Fourth	Walla Walla	59	6	.30
Do	Fourth	Muroc	44	6	.43
Do	Fourth	March Field I	55	5	.26
Do	Fourth	Tonopah	45	8	.67
Odd v. even releases	Second	Biggs	54		.85
1st v. second halves training	Second	Boise	154		.42
Low v. high altitude missions	Second	Ardmore[2]	69		.40
Do	Second	Biggs	39		.30
Do	Second	Pueblo	56		.35
Medium v. high altitude missions	Second	Biggs	54		.31

[1] Since a strictly odd-even pairing of missions coincided with systematic differences in type of missions, a random choice of missions was used in pairing scores.

[2] Only those missions with all strikes photographed are included in these data.

Although the reliability coefficients for odd and even missions covered a considerable range, practice bombing ACE appears to have had from low to moderate reliability.

Radar Bombing

In general, radar bombing in the Continental Air Forces was confined to operational training for B–29 crews. Training methods were designed to offer B–29 crews realistic practice in bombing industrial targets. Major emphasis was placed on the smooth coordination of bombardier, radar operator, navigator, and airplane commander. Since industrial targets were required for simulated combat training, there was obviously no actual bombing involved. Instead, scoring was accomplished either by a camera mounted on the underside of the fuselage (the more usual method) or by a ground radar installation in the vicinity of the target.

Several methods of evaluating radar bombing performance were used in the Second Air Force. One method of scoring, referred to as Photo Score (Victorville), required a series of photographs at specified intervals. The time of release in the series of photos was known. From the track of the airplane and other data such as altitude, winds, the theoretical point of impact was estimated. The Photo Score (Nadir) involved taking one

photograph at the moment of simulated release. By triangulation, the point immediately below the plane at the time of release was determined and from this information plus data such as altitude, winds, speed of the aircraft, the theoretical point of impact was estimated. The third scoring method involved the use of ground radar installations. These installations were able to tell the track, altitude and all other necessary information concerning the airplane at the time of simulated release to determine the amount of error in the bombing performance. There are relatively few data on which to base comparisons of the three types of scoring. Reliability coefficients for odd-even bomb runs were 0.10 for the first photo score (N of 44), 0.51 for the nadir score (N of 140) and 0.09 for the automatic scoring system (N of 51). It is believed that the ground radar method of scoring is potentially the most successful method.

Since each of the three scoring systems was used at one time or another with the crews being studied it was not possible in the Second Air Force

TABLE 7.5.—*Reliability of over-all radar terminal CE T-scores, odd-even missions*

16th AND 17th WINGS—SECOND AIR FORCE

Station	N	M, *1*	SD, *1*	M, *2*	SD, *2*	r_{tt}
Alamogordo	20	51 90	6.62	49.40	6.02	0.45
Alboquercue	23	48 10	4.78	50.02	6.16	.36
Clovis	36	49.69	4.67	49.28	3.78	.28
El Paso	25	49.06	3.48	50.14	7.54	.28
Great Bend	31	51.76	5.72	48.98	5.80	.52
Tucson	38	49 72	5.56	49.14	6.18	.50
Walker	46	49 16	4.99	49.68	5.35	.54
Combined (average by Fisher's z)	219	49.84	5.48	49.48	5.80	.44 [1].61

[1] Corrected by Spearman-Brown formula.

to get a large body of data in which just one method was used. The reliability coefficients reported in table 7.5 were obtained as follows: First, the average score for each crew on each mission was converted into a T-score based upon the distribution of all scores at the same station of the particular type involved. Then the T-scores for all three types of odd and even mission scores were averaged to obtain over-all odd and even mission radar terminal CE T-scores. Cases which did not have at least four scores of one type of scoring were discarded. Since T-scores were determined from separate distributions by stations for each type of scoring, all cases from different stations were combined directly.

All radar bombing accomplished at the Lead Crew School, Muroc Army Air Field, Fourth Air Force, was scored by camera. Both the Victorville and Nadir scoring procedures described above were used, no distinction being made in the analysis of results. Reliability coefficients computed from odd and even missions are presented in table 7.6. These data were not corrected for altitude. When data for the M class were corrected to 12,000 feet the reliability coefficient was reduced n 0.82 to 0.76.

A tendency toward bimodality noticed in the istributions of circular errors for the first class at Muroc AAF (approx ately 1,000 bombing

scores) suggested to research personnel the possibility that there were two independent factors operating in the radar circular error scores. One of these might be called "gross error," in which the target was misidentified; and the second was "normal aiming error." This observation led to further analysis of the Muroc data. On the basis of the distributions of the first class, it was decided that errors of 6,000 feet and over would be regarded as gross errors, resulting from failure to identify the proper target. It was hypothesized that these errors resulted from failure on the part of those persons with radar scopes (radar operator and navigator) to correctly interpret the blip on the scope. Errors less than 6,000 feet were believed to be largely influenced by the skill of the bombardiers in synchronizing the sight and introducing final corrections.

The reliability of normal aiming errors was estimated from odd-even mission scores. Data from gross errors were treated as percentages of the

TABLE 7.6.—*Odd-even mission reliability coefficients for measures of radar bombing proficiency*

MUROC AAF, FOURTH AIR FORCE

Measure	May		June		July		August		Total	
	N	r_{tt}	N	r_{tt}	N	r_{tt}	N	r_{tt}	N	r_{tt}
Radar ACE total	12	0.82	23	0.61	38	0.76	41	0.68	114	0.71
Radar ACE normal aiming	...	......	23	.65	38	.70	41	.45	102	.59
Percent gross errors	...	......	23	.20	38	.65	41	.61	102	.51

total runs for odd and even missions, since the actual magnitude of gross errors was irrelevant. These reliability coefficients are also presented in table 7.6. These two scores, ACE of normal aiming and percentage of gross errors, were correlated 0.10, 0.41 and 0.45 for the June, July and August classes. This degree of correlation must be regarded as moderate, considering the reliabilities of each. Comparison of the reliability coefficients of the various scores suggests that the presence of gross errors in total circular error did not significantly reduce the reliability of the raw scores. This indication that the two factors were not as independent as at first thought was substantiated by data from subsequent classes. Distributions of these later data did not show bimodality and were very nearly normal except for a long tail of extremely high error scores. This long tail of large error scores is probably a function of the bombing situation in which there is virtually no limit to the extent by which a target can be missed.

While the distinction between normal aiming and gross errors did not provide a great deal of insight into the nature of the function, the resulting analysis demonstrated that the gross error score was moderately reliable. This result is of practical value since it involves far less computation than does the usual average circular error score.

From the standpoint of reliability, radar average circular error scores are regarded as useful criterion measures.

Ratings

Rating scales for bombardiers were used extensively throughout the Continental Air Forces. In most cases, however, the amount of time required to develop rating methods precluded a thorough follow-up of their effectiveness. The succeeding paragraphs briefly describe a number of representative rating methods.

Efficiency ratings from Form 66-2 were available throughout the Continental Air Forces. These were ratings on a 5-point scale which were assigned at 3- month intervals by supervisory personnel. Since a single rating was submitted for each rating period, no direct estimate of the reliability of this measure could be made. However, at one station in the Second Air Force the reliability coefficient computed from pairs of successive ratings one month apart was 0.37 for 109 bombardiers.

An instructors' questionnaire for bombardier trainees was developed for use in the Third Air Force. The six items in the scale were: cooperative attitude, emotional stability, common sense, over-all bombardier proficiency, motivation and initiative, and desire to serve with trainee in combat. To minimize halo effect a non-numerical descriptive rating technique was employed. Neither item nor total scale reliabilities were computed.

A preliminary rating scale was used in the Fourth Air Force for the evaluation of B–24 bombardiers at three bases: Muroc, Walla Walla and March Field. The scale consisted of ten questions in response to which instructors were requested to rate (half above-average, half below-average) all of the students whom they had check-ridden. Except for the inclusion of an over-all anchor item, the questions referred to the same traits as did the ranked items listed in table 7.1. In order to minimize halo effect the questions were administered successively on separate sheets of paper, on which were typed the names of all students in the section. Since adequate samples of bombardiers rated more than once were not available, no estimate of item reliabilities was made.

B–29 bombardiers at the Lead Crew School, Muroc Army Air Field, Fourth Air Force were evaluated by a mission rating system (*see* table 7.2 for the scale items and weights). On every mission for which a bombardier instructor was aboard, students were rated on a 5-point scale on each of the 14 items. In addition, the instructors were required to make an over-all evaluation of the bombardier as a potential lead crew member. The odd-even reliabilities of the average score (the average of the ratings on the 14 items) and the over-all score are presented in table 7.7. Both scores show moderate reliability and are quite highly intercorrelated (0.70 on a sample of 94).

Two types of descriptive rating scales were used in the Second Air Force for evaluating B–29 bombardiers. The traits for which items were constructed were obtained from interviews with combat personnel, from manuals and directives for selection and training of lead crews, and from

intelligence reports and mission analyses. Scale A was administered at four stations: Clovis, Alamogordo, Tucson and Biggs; and Scale B was administered at another four stations: Albuquerque, Pyote, Walker and Great

TABLE 7.7.—*Reliabilities of ratings of bombardier proficiency*

LEAD CREW SCHOOL—FOURTH AIR FORCE

Measure	June		July		August		Total	
	N	r_{tt}	N	r_{tt}	N	r_{tt}	N	r_{tt}
Average rating	23	0.51	38	0.53	32	0.58	90	0.5[illegible]
Over-all rating	23	.82	38	.38	32	.39	93	.40

Bend. All student bombardiers included in the study were rated independently by two instructor observers. Distribution statistics are presented in table 7.8 and reliability coefficients for Scale A and B are presented in table 7.9.

TABLE 7.8.—*Distribution of bombardier ratings*

SCALES A AND B, SECOND AIR FORCE

Scale A Rating	N	*Scale B* Rating	N
2	0	4	0
3	6	6	7
4	11	8	19
5	22	10	21
6	35	12	28
7	51	14	41
8	56	16	30
9	72	18	35
10	33	20	60
11	23	22	41
12	5	24	19
13	4	26	13
14	2	28	10
		30	8
Total	320	Total	332
Mean	8.0	Mean	18.09
S.D.	2.1	S.D.	5.62
Number of items	3	Number of items	7

TABLE 7.9—*Reliability coefficients of total scale scores*

SCALES A AND B—SECOND AIR FORCE

Station	N	M, *1*	SD, *1*	M, *2*	SD, *2*	r_{tt}
Scale A:						
Alamogordo 3	23	7.43	1.66	8.57	1.38	0.44
Alamogordo 6	23	6.92	2.06	6.74	2.21	.88
Biggs	41	8.71	2.01	9.02	2.11	.84
Clovis	36	9.44	1.50	7.50	1.42	.59
Tucson	37	7.43	1.88	7.14	2.01	.78
Combined (average by Fisher's z)	160	8.14	2.05	7.85	2.07	.74
Scale B:						
Great Bend	14	2.66	0.83	2.52	0.49	0.53
Walker	46	2.31	.77	2.54	1.09	[illegible]
Albuquerque	49	2.78	.79	3.07	.76	.06
Pyote	33	2.47	.84	2.45	.93	.55
Combined (average by Fisher's z)	142	2.54	.82	2.63	.79	.46

In addition to the foregoing scales, a variety of local rating devices, too numerous to describe, were encountered at various training stations throughout the country. Whatever their form, these were usually *ad hoc* arrangements designed to meet special administrative requirements. The inadequate dispersion of ratings usually obtained made statistical analysis of these procedures not worth while.

Ground Trainer Grades

Among ground trainers for which adequate information was available, only the 7-A-3(A-5) trainer was designed strictly for bombardiers. The 7-A-3 trainer consisted essentially of a simulated bombardier's compartment of a B-17 or B-24. Equipment included a Norden bombsight, pilot direction indicator, automatic bombing computer, C2 and E6B computers, and intervalometer. A projector mounted near the top of the trainer housing projected a moving panoramic terrain on a screen beneath the bombardier's compartment, so that actual flight was rather closely simulated. The trainer was so designed that a number of objective scores were readily available. Of these scores circular error alone was investigated. Odd-even mission reliabilities computed for samples from three air forces are reported in table 7.10.

TABLE 7.10.—*Reliability coefficients for 7-A-3 trainer average circular error, estimated from correlations between odd and even missions*

Source of data	N	Number of missions	r
2d Air Force, Great Bend	54	8	0.40
2d Air Force, [illegible]	56	10	.70
4th Air Force, Walla Walla	56	3	.25
4th Air Force, Muroc	44	5	.48
4th Air Force, March Field I	43	7	.47

Scores from navigation and gunnery trainers were available at most stations. The dead-reckoning (G-1 navigation) trainer yielded an odd-even mission reliability of 0.47 for 542 bombardiers at one station in the Second Air Force. On the Jam Handy gunnery trainer the reliability coefficient obtained from the average intercorrelation of 3 scores (adjusted by the Spearman-Brown formula) was 0.41 for 77 Second Air Force bombardiers.

Academic Grades

Academic grades from written tests and class room performance were available throughout the Continental Air Force. The tests were notable mainly for their lack of uniformity in content and construction from station to station. Most frequently encountered were tests covering equipment, C-1 autopilot, theory of bombing, gunnery and communications. In the Second Air Force reliability coefficients estimated from scores on the first and second half of training were 0.75 for gunnery grades (N of 27), 0.55 for bombardment (N of 165), and 0.41 for communications grades (N of 54).

Flying Evaluation Board Action

The principal function of Flying Evaluation Boards (FEB's) was to investigate and appraise aircrew officers who had refused to fly or who were charged with some type of incompetence. The major difficulty encountered in interpreting FEB decisions as potential criteria was the lack of standardization in methods of investigation and evaluation. Table 7.11 presents the frequencies with which various categories of reasons were cited as principal cause for board action, for samples of evaluated bombardiers from the Second, Third and Fourth Air Forces.

TABLE 7.11.—*Frequencies of principal reasons for referral of bombardiers to Flying Evaluation Boards*

Principal reason	Second Air Force	Third Air Force	Fourth Air Force
Personality defect (mainly fear)	36	31	10
Lack of proficiency	12	2	3
Physical disability	10	0	3
Total	58	33	16

Intercorrelations

Of the intercorrelations shown in table 7.12 among the criteria available for B–24 bombardiers in the Fourth Air Force only one, between G-1 trainer rating and average academic grade is significant at the one precent level. Since at least some of the measures are known to have had moderate

TABLE 7.12.—*Intercorrelations among measures of proficiency of bombardiers (B–24)*

THREE FOURTH AIR FORCE BASES

Proficiency measures	1	2	3	4	5
1. Over-all rating		0.15	−0.01	0.12	0.29
2. Average circular error	146		.05	.06	.05
3. 7-A 3 circular error	142	251		.06	−.27
4. Average academic grade	144	306	251		−.45
5. G-1 trainer rating	41	41	41	41	

NOTE—Numbers of cases appear below the diagonal. Signs of coefficients have been changed to indicate positive association of good performance.

reliabilities some independence of skills represented in the scores can be presumed.

As has been stated earlier, the 2 rating scores for B–29 bombardiers at the Lead Crew School correlated 0.70 with each other for a sample of 94 cases. As shown in table 7.13 the rating scores showed little relationship to the bombing scores.

TABLE 7.13.—*Intercorrelations among criteria available for B–29 bombardiers*

LEAD CREW SCHOOL[1]

	1	2	3	4	5
ACE total score	......	0.74	0.87	−0.01	0.09
ACE normal aiming	......	......	.45	.18	.17
Per cent gross errors	......	......	......	−.06	.07
Average rating	......	......	......	......	.70
Over-all rating	......	......	......	......	......

[1] N = 94.

This lack of relationship between radar error scores and ratings is confirmed by the data from the Second Air Force. In the case of 125 cases with Scale A, the correlation between the radar scores and ratings was 0.13 while for Scale B on 86 cases the coefficient was zero. However, in the case of B-17 bombardiers the relationship between practice bombing average circular error scores and ratings was considerably higher as the results in table 7.14 indicate. That this relationship is not due solely to differences in aircraft is attested by the correlation of 0.58 between Rating Scale A total scores and practice bombing ACE for 160 B-29 bombardiers. The higher relationship in the case of circular errors for practice bombs is due in part to the fact that the raters were more likely to know the practice bombing score. Among the raters utilized in these studies the flight surgeon alone was unaware of the bombardier's average circular error, and this coefficient was markedly lower than the others. The radar bombing score was much less likely to be known to the raters for a number of reasons. No record of radar scores was available to the raters at the time the ratings were made. In addition, the radar score was not usually regarded as solely a measure of bombardier proficiency, being more often regarded as a reflection of the bombing skill of several members of the crew. Thus, even if raters knew the radar bombing score, it was not as likely to affect the rating as knowledge of the average circular error of practice bombing.

TABLE 7.14.—*Correlations between average circular error (practice bombing) and various rating scores*

B-17 AND B-24 BOMBARDIERS—SECOND AIR FORCE

Type of rating	N	r
General aircrew ability (AFTRC Form 2)	111	0.[illegible]
Efficiency rating (Form 66-2)	133	[illegible]
Rank in flight	640	[illegible]
Interview rating	52	.[illegible]
Crew rating	52	.28
Flight surgeon's rating	52	.06
Self-rating	52	.15

A low but statistically significant correlation between average ground school grades and both radar and practice bombing scores was found for B-29 bombardiers in the Second Air Force. Both coefficients were biserials between ACE and "high-low" classification on the basis of the average of all ground school grades. The coefficient was 0.21 in both cases on a sample of 159 students for practice bombing and 124 for radar bombing. Both coefficients are significant at the 5 percent level. However, practice bombing ACE showed no relationship to ground school average grade and test scores in the case of 112 B-17 bombardiers, as is shown by the results in table 7.15.

Rating scale scores for Scale A correlated 0.1[illegible] with ground school grades for a sample of 160 B-29 bombardiers. This biserial correlation coefficient, computed as described for bombing scores above, is significant at the 5 percent level.

In the Third Air Force, the bombardier questionnaire score, described earlier, correlated 0.83 with the rankings of 130 bombardiers by the same instructors making the entries on the questionnaire.

In general, the degree of correlation observed between different criteria suggests tentatively that: (1) Rating scores of various sorts had little or nothing in common with radar bombing scores, (2) rating scores were moderately related to practice bombing scores, (3) bombing scores of both types appeared to have some slight correlation with ground school performance and finally, (4) rating scores did not show any such relationship.

TABLE 7.15.—*Correlations between average circular error (practice bombing) and ground school measures*

B-17 BOMBARDIERS—SECOND AIR FORCE

Measure	N	r
Entrance test	52	0.13
First phase test	52	−.16
Second phase test	52	.02
C-1 autopilot examination	52	.05
Ground school average grade	112	.06

Evaluation of Criteria

Evaluated in terms of their reliability, practicability and apparent relevancy to the ultimate bombardier criterion, the criteria described in the foregoing paragraphs appear to be promising. As a whole they cover pretty well the principal duties of the bombardier in operational training. Their reliabilities, although far from high, are satisfactory minimum values. They create few major administrative problems in use, and they are available throughout the Continental Air Forces. However, in each case there are certain serious objections to their use as criteria of bombardier proficiency.

Average Circular Error

The average bombing error is in many ways the most impressive criterion available. It has an inherent validity not possessed by any other measure of crew effectiveness. However, its full value has often been reduced by a number of factors. The most serious difficulty met in using average circular error as a criterion of bombardier proficiency in operational training was the fact that it was probably influenced nearly as much by pilot proficiency as by bombardier proficiency. In this sense, average circular error may be more a measure of crew proficiency than a measure of proficiency of any one crew member. In fact, it is likely that the fact that the same pilot and crew fly with the bombardier on his bombing missions is responsible for the higher degree of reliability found for average circular error in operational training as compared with average circular erro in Bombardier Schools in the AAF Training Command. In the latter case the differences in pilot proficiency tended to reduce the reliability of circular error scored since different pilots flew with a bombardier on different missions.

One of the major sources of distortion of average circular error has been clerical inefficiency. At one station in the Third Air Force approximately 70 percent of the bombing records inspected contained clerical errors. Errors were found consistently in such operations as the recording of altitude, range direction, and raw circular error, in the conversion of raw error to 12,000 feet, in the product of the number of bombs dropped and converted circular error, and in the cumulating of converted circular error.

Another important source of distortion was the treatment of gross errors. In practice bombing, bombs dropped off the range were neither scored nor recorded. However, if a bombardier could (or cared to) estimate such an error, his estimate was recorded. In radar bombing, when scoring was accomplished by camera, "strikes" which could not be identified on a master photograph of the target were ignored even when they obviously represented gross errors. Thus, a great majority of the strikes reported as "unscorable" were probably very large errors, for which no record was available, and an important segment of the data was overlooked. Should further analysis establish that "percentage of gross errors" is satisfactorily reliable, it might be reasonable to record that datum in addition to ACE and to analyze bombing error routinely in terms of both scores. Normal aiming errors in that case would be treated as magnitudes, and gross errors as relative frequencies.

In addition to the above-named factors, many other distorting influences can be cited. Many of these are so apparent that comment is unnecessary. Standardization of training requirements from trainee to trainee would do much toward eliminating a host of factors which attenuate the reliability of circular error. Improved scoring procedures, such as sonic recording of practice bombing errors and ground radar scoring for radar bombing, will doubtless reduce other important sources of error.

Ratings

Insufficient evidence makes it extremely difficult to evaluate the rating scales described above. For example, it is not now possible to decide with the evidence at hand whether the Second Air Force descriptive scales, which are administered at the end of training, were more or less efficient than the less refined Muroc scale which was administered mission-by-mission. On the whole, the rating devices appear to have had moderate reliability, and their designs appear to encompass adequately the breadth of the bombardier's essential skills and personality traits. Improvement of results with such scales probably rests as much on the future indoctrination of instructors in good rating practices as on refinements in the scales themselves. However, the chief objection to the use of ratings is their inherent susceptibility to subjective errors. The moderate degree of reliability obtained may be largely a reflection of group consistency of prejudices or of common reputation, rather than a reflection of agreement in independent observations.

Ground Training Scores

Clerical errors and imperfect standardization of content and procedures of administration constituted the principal sources of attenuation in ground trainer scores. It is also likely that an increase in the amount of time allotted to ground trainer instruction would yield better data. Since all of these shortcomings are remediable, it should be possible to obtain scores that are objective as well as reliable.

Academic grades suffered from a variety of difficulties. Inexpert construction of tests, constant modification to include technical innovations, inadequate standardization, clerical errors, and poorly motivated students all combined to make them relatively meaningless. A number of remedies are obvious and practicable. Perhaps the clearest need here is for a central test-construction agency whose responsibility would be the design and continual modification of a series of standard tests.

Flying Evaluation Board Reports

Re-evaluation of bombardiers by action of Flying Evaluation Boards did not occur sufficiently often for this type of criterion to be of much practical value in validation studies.

The use of Flying Evaluation Board records for validation purposes also appeared to suffer from lack of standardization. Uniform procedures, uniform rules of evidence and experienced board members who are fully grounded in the rules and procedures are essential requirements, if the board decisions are to become a satisfactory criterion.

VALIDATION STUDIES

Average Circular Error

Practice Bombing

With average circular error as the criterion, validity coefficients for stanines were computed for 437 B–17 bombardiers in the Second Air Force. The results of this computation presented in table 7.16 indicate that the relationship between the bombardier stanine and circular error was not statistically different from zero.

TABLE 7.16.—*Validities of the bombardier stanine for average circular error*

B-17 AND B-24 BOMBARDIERS—SECOND AIR FORCE

Station	N	B stanine, 1		Circular error, 2		r_{12}
		M	SD	M	SD	
Ardmore	155	6.56	1.14	240	39	-0.0[illegible]
Biggs	52	6.81	1.0	258	49	-.0[illegible]
Boise	212	6.12	1.5	318	67	.0[illegible]
Pueblo	54	6.64	1.1	222	34	.1[illegible]
Total	473	.....	.....	.....	.....	.0[illegible]

Radar Bombing

The validity coefficients for the stanines for 70 B–29 bombardiers in the Second Air Force and 84 B–29 bombardiers at Muroc in the Fourth Air

Force are presented in table 7.17. The coefficients given in the table represent averages (by Fisher's z) of separate coefficients for each class and station. As was the case with practice ACE, the stanines of bombardiers were not significantly related to any of the radar bombing scores.

TABLE 7.17—*Correlations between stanines of bombardiers and radar bombing scores*

Type of score	Number of Cases	Radar bombing score		Bombardier stanine		Navigator stanine		Pilot stanine		r_B	r_N	r_P
		Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.			
ACE T-scores, 2 AF	70	50.3	4.56	6.54	1.13	5.87	1.15	.....	.....	−0.05	0.04	.....
Total ACE, 4 AF	84	a	a	6.51	1.46	5.41	1.19	5.40	1.89	.07	.09	0.00
Normal aiming ACE, 4 AF	84	a	a	6.51	1.45	6.41	1.19	5.40	1.89	−.12	−.07	−.06
Per cent gross errors, 4 AF	84	a	a	6.51	1.46	6.41	1.19	5.40	1.89	.13	.13	.17

Ratings

Table 7.18 shows the bombardier stanine validity coefficients obtained for various kinds of ratings of B–17 bombardiers in the Second Air Force. None of the coefficients is statistically significant except that obtained for flight surgeon's rating, which is negative, and crew rating which is positive. The small samples make these latter results of doubtful importance. Descriptions of the different ratings used are given in an earlier section under *Criteria*.

TABLE 7.18.—*Validities of bombardier stanine for various types of ratings*

B–17 AND B–24 BOMBARDIERS—SECOND AIR FORCE

Rating device	N	Stanine, 1		Rating device, 2		
		M	SD	M	SD	r_{12}
Efficiency rating (66-2)	105	6.41	1.25	44.80	11.00	−0.05
Instructors rankings	178	6.51	1.17	52.40	15.10	.64
Technical aircrew skill	53	6.02	1.27	51.60	8.90	−.04
Officer quality rating	53	6.02	1.27	51.60	9.40	.23
General aircrew ability	53	6.02	1.27	51.60	7.90	.13
Interview rating	50	6.80	1.16	5.64	3.00	.01
Flight surgeon's rating	50	6.80	1.16	1.88	.60	−.29
Self-rating	50	6.80	1.16	2.78	1.10	−.01
Crew rating	50	6.80	1.16	6.30	3.00	.26
Attitude toward work	212	6.12	1.50	2.26	.50	.11

In table 7.19 are presented validity coefficients for the three stanines of B–29 bombardiers in the Second and Fourth Air Forces, using rating scale scores as criteria. These scales are described earlier in the chapter. None of the coefficients is significantly different from zero.

In the Third Air Force, the bombardier, navigator and pilot stanines of 133 B–29 bombardiers correlated 0.02, —0.11, and —0.01 with ratings on the bombardier questionnaire described under *Criteria* above.

Ground Trainer Grades

Two studies were conducted to determine the extent to which the stanines of bombardiers predicted their performance in the 7-A-3 bomb

trainer. In the Second Air Force (Great Bend), the correlation coefficients between 7-A-3 circular error and the bombardier and navigator stanines for 32 B–29 bombardiers were —0.02 and 0.16 respectively. The predictive efficiency for trainer performance is somewhat greater in the results from the Third Air Force. The validity coefficients were 0.13 for the bombardier, 0.23 for the navigator and 0.01 for the pilot stanine on a sample of 49 bombardiers at Barksdale Field. None of these coefficients is significant at the 5 percent level.

TABLE 7.19.—*Correlations between instructors' ratings and stanines of bombardiers*

B–29 BOMBARDIERS, SECOND AND FOURTH AIR FORCES

Stanine	N	Stanine, 1		Rating, 2		r_{12}
		Mean	SD	Mean	SD	
Scale A:						
B	148	6.66	1.22	16.34	3.56	†0.03
N	148	6.03	1.27	16.34	3.56	†.02
P	148	5.48	1.57	16.34	3.56	–.06
Scale B:						
B	159	6.59	1.20	37.26	9.36	–.04
N	159	5.81	1.12	37.26	9.36	–.04
P	159	5.31	1.60	37.26	9.36	.07
Muroc average rating						
B	79	6.46	1.48	......	......	–.16
N	78	6.43	1.23	......	......	.01
P	78	5.44	1.92	......	......	.01
Muroc over-all rating:						
B	79	6.46	1.48	......	......	.09
N	78	6.43	1.23	......	......	.09
P	78	5.44	1.92	......	......	.03

Academic Grades

Table 7.20 shows the bombardier stanine validity coefficients for various types of academic grades and ground training criteria. Attention is called to the fact that the bombardier stanine is significantly related to three of these measures: academic average, ground school average, and bombardier entrance test.

TABLE 7.20.—*Validities of the bombardier stanine for ground trainer criteria*

B–17, B–24 AND B–29 BOMBARDIERS—SECOND AIR FORCE

Criterion	N	B stanine, 1		Criterion, 2		r_{12}
		M	SD	M	SD	
Academic average	212	6.12	1.50	82.30	4.60	*0.16
Ground school grades	80	6.53	1.12	45.90	10.70	*.23
Gunnery grade	49	6.90	1.18	81.90	3.90	.08
Bombardment grade	49	6.90	1.18	85.60	6.10	.16
Bombardier proficiency test	56	6.64	1.08	78.00	9.30	.17
Bombardier entrance test	56	6.64	1.08	83.80	10.10	*.30
C-1 autopilot examination	56	6.64	1.08	86.90	11.10	.01
Bombardment grade (B-29)	32	6.84	1.18	91.81	2.66	.01

In the 16th Wing of the Second Air Force, the B–29 bombardiers of five classes were divided into "upper" and "lower" groups on the basis of their average ground school grades. The biserial correlation between the bombardier stanines and this classification are presented in table 7.21. These results are quite different from those obtained with B–17 and B–24 bombardiers. Whether this is brought about by the fact that B–29 training

was newer and less well standardized or whether other factors are responsible is not known.

Table 7.21.—Biserial correlations between bombardier stanines and ground school grades

B-29 BOMBARDIERS—SECOND AIR FORCE

Station	N_t	M_t	SD_t	M_o	M_f	P_o	r_{bis}
Alamogordo 3	21	6.86	1.08	7.00	6.75	0.43	0.16
Alamogordo 6	19	6.79	1.10	6.58	7.14	.63	−.32
Biggs Field	38	6.47	1.27	6.52	6.38	.66	.07
Clovis	33	6.48	1.44	6.24	6.75	.57	−.22
Tucson	37	6.84	1.05	6.84	6.83	.51	.01
Combined	148	6.66	1.23	6.60	6.74	.55	[1]−.06

[1] Combined by means of Fisher's z technique.

Flying Evaluation Board Reports

In an effort to predict elimination of bombardiers by Second Air Force Flying Evaluation Boards, the bombardier stanines of 123 bombardiers relieved from flying duties by FEB's were compared with those of 123 bombardiers who (1) were never referred to FEB's, (2) were in operational training at the time of FEB action and (3) were tested at a classification center at approximately the same time as the eliminated bombardiers. The results, which appear in table 7.22, indicate that the stanines of bombardiers eliminated by FEB action are significantly lower than those of a randomly selected control group. The stanines of 15 Third Air Force bombardiers who were eliminated by FEB action were slightly (but not significantly) higher than those of a control group of the same size.

Table 7.22.—Bombardier stanines of bombardier FEB eliminees compared with those of control samples of non-eliminees

B-17 AND B-24 BOMBARDIERS—SECOND AND THIRD AIR FORCES

Air force	N	FEB cases		Controls		Critical ratio
		M	SD	M	SD	
Second	123	4.71	2.16	5.59	1.91	**3.14
Third	15	5.33	2.36	4.87	1.81	[1].30

[1] Direction of difference contrary to expectation.

SUMMARY

Most of the time spent in research on bombardier training was spent in the discovery and evaluation of measures of proficiency. Bombing scores, trainer grades, ratings, rating scale scores and academic grades were rather thoroughly explored. With the exception of routine ratings and academic grades, all of these were found to have a degree of reliability which, though low, was sufficient for their use as proficiency criteria. Because of susceptibility to subjective errors, rating scale scores were not considered very useful criteria in themselves. Ground trainer grades were based on procedures which varied too much from one station to another. The measure

that has the highest inherent validity, the bombing score, did not, in actual practice, sufficiently reflect the prowess of the bombardier to be completely satisfactory. Thus, no one measure was found which could be considered a completely satisfactory criterion of bombardier proficiency.

The stanines of the bombardiers in the samples studied showed little or no correlation with most of the criteria studied. There is evidence that the bombardier stanine had some relationship to certain measures of proficiency in academic subjects. The general inadequacy of the available criteria makes this lack of predictive efficiency on the part of the bombardier stanine difficult to interpret. It is believed that effective validation of any device for prediction of bombardier proficiency in operational training will have to wait upon the development of measures of bombardier proficiency that are more adequate than any tried so far.

CHAPTER EIGHT

Flight Engineer

ANALYSIS OF DUTIES AND SPECIFICATIONS

Job Analysis

The flight engineer was a relative newcomer among the officers of the aircrew. The job, of course, was in existence for some time before the position was formally recognized. It had its beginnings in the division of labor among enlisted crew members of early two and four engine aircraft. One crew member was charged with responsibility for certain preflight inspections of engines and for supervision of the fuel and oil systems. With the increasing complexity of bombardment aircraft the importance of the flight engineer increased enormously. Not only was importance increased, but the duties became correspondingly difficult and complex, until in the heavy bombardment crew, the necessary skills and knowledge required organized courses in special schools. With the advent of the B–29 aircraft, training became still more extended and included the additional step of transition training of airplane commander, copilot and flight engineer as a team prior to operational training of the total crew. At this point, the flight engineer was made a commissioned officer on a par with copilot, navigator and bombardier.

For a number of reasons no detailed analysis of the duties of the flight engineer was carried out in the Continental Air Forces. Training in the basic skills and routine handling of the job is done in the AAF Training Command so that such studies can probably best be accomplished there. Then too, a Psychological Research Project was set up in connection with that training with specific responsibility for developing selection procedures and methods of evaluating flight engineer proficiency. Effort in the Continental Air Forces was therefore concentrated upon studies of those aspects of the job that were peculiar to operational training and those especially important in the effective performance of the whole crew.

In operational training, the flight engineer learned to apply his knowledge of cruise control, of fuel consumption, and of all systems of the aircraft to practical problems in simulated combat operations. He had to become so familiar with the power plant, mechanical and electrical systems of his plane that this knowledge was almost second nature to him. He needed to know the location of all tools and emergency equipment and be able to make minor repairs in flight. In B–29 crews especially, he had to be able at all times to estimate accurately the load the plane was carrying and its center of gravity. He had to be able to make accurate estimates of the

changes in fuel consumption and load caused by any special conditions of engine functioning or irregularities in flight. It was his especial responsibility in operational training to develop an effective team to carry out inspection and continuous observation of instruments and engines. Like the pilot and copilot, the flight engineer flew whenever his plane flew, regardless of the purpose or nature of the mission. He thus received the full 120 to 126 hours of flying time specified for combat crews in operational training.

Job Specifications

Certain qualities seemed to be more important for success as a flight engineer than others. First of all, regardless of the situation in which he found himself, the flight engineer had to be able to carry out—accurately, systematically, and quickly—the arithmetic computations necessary in cruise control and load determination. Second, he had to be a good "estimator." If one of the engines started to backfire and had to be run "auto-rich" its fuel consumption changed. Skill in being able to judge the amount of such change was important. The B–29 flight engineer also had to have sufficient of the qualities of leadership to be able to organize the enlisted crew members into an efficient team to keep engines and instruments under continuous observation at all times. Finally, in common with other members of the crew, he had to be cooperative; for he had to coordinate his activities with pilot, copilot and navigator.

CRITERIA OF PROFICIENCY

General Observations

Observations of training records and conversations with instructors and officers in charge of training indicated that there were no routinely available records that would be of much value in determining flight engineer proficiency. The only more or less objective record which seemed likely to provide such information was the flight engineer's log. These proved to be very disappointing under current practices. Evaluation of these logs was neither systematic nor standardized although the form was prescribed. Different instructors at the same station often used different procedures in evaluation. Discrepancies between estimated and actual fuel remaining at the end of a flight were considered as a possible source of information on proficiency. However, no such records were ordinarily kept and any attempt to use such a criterion would have required the institution of a special report. In the absence of more objective criteria research personnel were for the most part forced to fall back upon subjective judgments in their evaluation of flight engineer proficiency in operational training.

Ratings by Instructors

In an attempt to develop procedures for the evaluation of crews as potential lead crew material, two descriptive rating scales were constructed

in the Second Air Force. These scales required five-point ratings on each of several traits for each aircrew officer, with some similar ratings for the crew as a whole. Scale A was administered at four CCTS stations (VH) while Scale B was administered at two OTU (VH) stations and two other CCTS (VH) stations. Each scale was administered to a total of more than 150 crews. Wherever possible two independent ratings were obtained on each crew member on each scale item. More detailed information on these ratings is given in chapter 11 on Lead Crew Proficiency, and chapter 4, Airplane Commander.

The two rating scales differed greatly in their items for flight engineers. Scale A contained three such items. The qualities to be rated were: Ability to meet unexpected changes in flight plan; ability to lead and supervise other crew members; and general lead crew proficiency. Scale B included seven items for rating flight engineers involving these qualities: Suitability for lead crew assignment; adequacy of preflight procedures; interest in the job; proficiency in cruise control procedures; ability to plan ahead; cooperation with the rest of the crew; and over-all proficiency. In table 8.1 are given separate distributions of ratings on each scale obtained with

TABLE 8.1.—*Distribution of flight engineer ratings*

SCALES A AND B—SECOND AIR FORCE

Ratings on single items

Rating	Scale A, item number			Scale B, item number						
	1	2	3	1	2	3	4	5	6	7
1	48	54	47	16	26	19	13	22	32	25
2	99	146	48	65	85	61	59	77	68	68
3	117	82	153	96	85	99	95	87	71	108
4	44	34	58	47	33	36	31	38	27	31
5	12	4	14	14	3	15	8	8	6	12

Ratings on all items combined

Scale A—3 Items

Rating	*N*
3	23
4	18
5	21
6	47
7	43
8	64
9	29
10	24
11	20
12	23
13	5
14	2
15	1
Total	320
Mean	7.65
S.D.	2.59

Scale B—7 Items

Rating	*N*
6–7	11
8	4
10	14
12	28
14	41
16	35
18	46
20	55
22	18
24	23
26	10
28	10
30	4
32	4
34	3
Total	306
Mean	18.39
S.D.	5.64

each flight engineer item and distributions of the combined ratings for all such items. In table 8.2 are shown the reliabilities of the total ratings expressed by correlations between the pairs of independent ratings of the same flight engineers. The correlations were first computed separately for each administrative unit—CCTS class or section and OTU squadron. These coefficients were then converted to the corresponding z values, averaged, and the resulting average z converted to the corresponding r. In this manner station tendencies to rate high or low do not contribute to the observed reliability. The reliability of 0.60 and the wide range of scores obtained with both scales are encouraging.

TABLE 8.2.—*Correlations between independent ratings of flight engineers*

SCALES A AND B—SECOND AIR FORCE

Station	Number of cases	Rater, 1		Rater, 2		r
		Mean	S.D.	Mean	S.D.	
Scale A:						
Alamogordo 3	23	2.13	0.85	2.19	0.92	[illegible]
Alamogordo 6	23	1.98	.58	2.22	.78	[illegible]
Biggs	41	2.11	.55	2.53	.39	[illegible]
Clovis	36	2.69	1.13	2.43	.91	[illegible]
Tucson	36	3.15	.65	3.01	.77	[illegible]
All stations (average by Fisher's z)	160	2.47	.90	2.52	.81	[illegible]
Scale B:						
Walker	32	2.55	0.60	2.29	0.49	[illegible]
Pyote	34	2.50	.76	2.50	.97	[illegible]
Albuquerque	50	3.23	.84	3.79	.80	[illegible]
All stations (average by Fisher's z)	116	2.86	.85	2.57	.82	[illegible]

Proficiency Check List

A rather different approach to the problem of estimating crew proficiency is found in the Crew Proficiency Check—VHB developed in the Third Air Force. With this instrument, an instructor observing a mission checked one of various alternatives under each of more than 50 items. One of these items referred specifically to the flight engineer. In this item the instructors were asked to check whether the flight engineers made use of a flight plan and gave data to the pilot, or whether they merely recorded the plane performance in their logs. This item of the Crew Proficiency Check was used with 120 flight engineers. Active participation was noted for 35, and passive recording of data for 75 for these flight engineers. In 10 cases the instructors felt that the behavior of the flight engineer did not fit into either category. No reliabilities are available for ratings on the Crew Proficiency Check.

Ground School Grades

Still another criterion of proficiency in operational training was the performance of the individual in his ground school classes. This performance of the individual was sometimes recorded in the form of separate scores on one or more written proficiency tests, or it represented the grade given for a whole course. The latter in turn may have been obtained by averaging

a number of test scores or it may have been an entirely subjective estimate by the instructor. Procedure varied greatly from station to station in a single Air Force, and probably varied still more from one Air Force to another. However, since many of the duties of the flight engineer and also of the navigator were of the nature of "paper work" there is perhaps better reason to use performance in ground school as a criterion than in the case of other members of the aircrew.

For two classes of flight engineers at Ardmore, the correlations between Phase II and Phase III average ground school grades were 0.54 (N of 56) and 0.82 (N of 55). Grades for these phases may be considered as first half and second half division of the training and the correlation between them is to a certain extent a measure of their reliability.

The correlation between grades on the over-all final examination in Engineering and on the final Cruise Control examination for the 4-21 class of B-29 engineers at Albuquerque was 0.10. The number of cases is small, but there seems to be little agreement between achievement in the course as a whole and knowledge of cruise control. For this same sample, the correlations between rating scale B total scores (raters 1 and 2) and the two examinations were virtually zero.

In the 16th Wing of the Second Air Force, the trainees in five classes were divided into "high" or "low" groups on the basis of their average ground school grades. Biserial correlation coefficients between this "high-low" criterion and Rating Scale A total scores are presented in table 8.3. The combined coefficient indicates a low but statistically significant (one percent level) relationship between these two criteria.

TABLE 8.3.—*Correlations between rating scale A total scores and ground school grades*

SECOND AIR FORCE

Station	N_t	P_u	M_u	M_l	SD_t	r_{bis}
Alamogordo 3	23	0.32	12.66	13.04	4.74	0.05
Alamogordo 6	23	.65	11.84	14.24	3.42	.42
Biggs Field	41	.54	13.78	13.96	2.52	.04
Clovis	35	.46	15.19	19.50	5.66	.48
Tucson	37	.49	18.40	19.54	3.98	.18
Combined	159	.52	14.54	16.63	4.83	**.23

VALIDATION STUDIES

Ratings

In table 8.4 are given the correlations between the stanines of the flight engineers and the total ratings, averaging the scores of two raters for the flight engineer items of Scales A and B. These coefficients were computed in two ways. First, separate coefficients were computed for each CCTS class or OTU squadron. A coefficient was then obtained for the total group by Fisher's weighted z method. Second, all cases on each scale were combined in a single group and a coefficient obtained in the usual way. Only the validity of the navigator stanine against the rating c ined on Rating

Scale A is markedly different from zero, this coefficient being significant at the one percent level. This is not corroborated by the data from the 17th Wing (Scale B). In general, the results indicate that, with this exception, the stanines did not predict the judgments of instructors regarding the proficiency of flight engineers.

TABLE 8.4.—*Correlations between stanines and rating scale scores (raters 1 and 2)*

SECOND AIR FORCE

Scale	Stanine	N	Stanine, 1		Ratings, 2		r_{12}	r_{12}[1]
			Mean	SD	Mean	SD		
A	B	129	6.24	1.04	15.40	4.88	0.08	0.12
	N	129	6.12	1.62	15.40	4.88	.22	-.28
P plus credit		129	6.57	1.61	15.46	4.88	.07	.10
B	B	94	6.19	1.87	38.06	10.56	.13	.04
	N	94	6.30	1.75	38.06	10.56	.09	.07
	P	94	6.63	1.69	38.06	10.56	.02	-.06

[1] Obtained by averaging coefficients for each class or squadron by means of Fisher's z technique.

It is of interest to note that the addition of another rater did not in this case appreciably affect validity of the pilot stanine. With a single rater the correlation between rating scores and pilot stanine was 0.06 and 0.03 for Scales A and B respectively.

Ground School Grades

The correlation between the augmented pilot stanines of flight engineers and their ground school grades expressed as "upper" and "lower half" of the class was 0.19. These data were obtained from four CCTS (VH) classes in the 16th Wing and the coefficient given is the weighted z average equivalent of the coefficients for the four separate classes. The combined validity coefficient for a total of 129 cases is significant at the 5 percent level.

The correlations between the pilot and navigator stanines of the flight engineers in the 8-21 class at Albuquerque and their grades in the over-all final examination in Engineering and final Cruise Control examination are given in table 8.5. Just as a matter of interest, the multiple correlation between navigator stanines and the combination of the two examinations was computed. Not much reliance can be placed on the high value of 0.56 obtained, since the number of cases is so small.

TABLE 8.5.—*Correlation between stanines and test scores*

B-29 CREWS—SECOND AIR FORCE

	Number of cases	Stanine, 1		Test Score, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Validity against pilot stanine:						
Over-all final test	32	6.31	1.45	93.5	4.70	-0.07
Cruise control test	32	6.25	1.41	89.9	3.25	.06
Validity against navigator stanine:						
Over-all final test	32	5.94	1.84	93.5	4.70	*.35
Cruise control test	32	5.97	1.79	89.9	3.25	**.48

SUMMARY

In view of the small amount of data obtained by the time of cessation of hostilities, it is difficult to make an evaluation. The negative results in table 8.4 could be explained in at least three ways: 1) The qualities expressed by the pilot stanine are not important for operational training of flight engineers; 2) The well-known sources of error in subjective ratings make them inadequate as a criterion of proficiency; 3) The narrowing of range of critical abilities by successive selections coupled with a great deal of variation in amount and type of previous training may make the small individual differences in ability actually present a relatively small factor in the observed variations in performance. Much additional study will be required before definite conclusions can be made as to validities of stanines or classification tests against flight data. However, it seems likely that the navigator, and to a less extent the pilot and bombardier, stanines of flight engineers do predict their success in ground school courses in operational training.

CHAPTER NINE

Radar Observer

ANALYSIS OF DUTIES AND SPECIFICATIONS

Job Analysis

As with the position of flight engineer, that of radar observer was a recent addition to the specialties of the combat crew. In operational training, in the Continental Air Forces, radar observers were included only on VHB (B–29, B–32) crews. On such crews, their main role was in navigation and in the bombing of targets obscured by clouds or darkness, etc. In both navigation and bombing the radar observer, though important, was only one of a cooperating team; and the success of the operation was the joint outcome of the activities of all team members. In navigation the team included the navigator, radar observer, airplane commander and flight engineer. In bombing the team consisted of bombardier, radar observer, airplane commander and navigator.

Obviously, the fundamental job of the radar observer was the handling of the radar equipment. This included the usual preflight checks, getting the set in operation and tuning, minor repairs in flight and interpretation of data from the sets for use in the desired operations. By identifying the patterns appearing on the "scope," the radar observer could make position fixes in navigation regardless of visibility. Range and bearing of all check points could also be obtained from the instrument. Accurate ground speeds, tracks and winds could be determined for use in pilotage and DR navigation. In bombing, the information for setting up and synchronizing the bomb sight was provided by the radar observer when the target was not visible to the bombardier.

Since in each case the radar observer was a member of a team cooperating in the operation, how effectively the information obtained by radar was presented to the other officers concerned was as important as the skill of the radar observer in handling the equipment itself. In navigation the radar observer and navigator worked independently, but each checked the other's work against his own. In this manner an accurate check was obtained of most of the computations, of the calculations of position and of the various values to be used in setting up the bomb-sight in the radar bombing operation. During the bombing operation, the radar observer provided the bombardier with information to set up and synchronize the bomb sight and then issued corrections on the basis of the evidence from the radar

scope. Then, if the target appeared in view at any time, the bombardier was prepared to switch immediately to visual bombing. If such was impossible, release was made on the basis of the radar settings for the usual radar bombing run.

The detailed description of the job of the radar observer in handling the equipment and in making the necessary adjustments and computations in navigation and bombing is beyond the scope of this report. Such can be obtained from a study of appropriate manuals and from Aviation Psychology Research Report Number 12: Psychological Research on Radar Observer Training.

Job Specifications

Successful handling of radar equipment undoubtedly required a high degree of ability to identify ambiguous and relatively unstructured visual patterns and to correlate a series of such patterns with features on maps of the terrain being covered. In the actual manipulation of the equipment, it was particularly important that the radar observer be skillful in adjusting various electrical controls so as to give maximum definition and clarity to the image. The ability to make fine manipulations under visual control was probably of primary importance here. Knowledge of the fundamental principles of radar operation and mastery of the details of the equipment used were necessary for in-flight repairs and adjustments and for the interpretation of any unusual features of operation of the equipment. The necessity of operating as a member of a team made ability to cooperate in a small group extremely important. A willingness to check all observations against reports of other crew members and ability to discover reasons for any disagreements encountered were essential to success in this specialty in the operational training of combat crews.

CRITERIA OF PROFICIENCY

General Observations

In addition to the usual difficulties brought about by differences in training procedures, differences in equipment and differences in records maintained, studies of proficiency of radar observers had to meet the difficulty that officers of this specialty were frequently not available for assignment to crews until all or a part of operational training had been completed. Furthermore, methods of evaluating radar bombing performance were in process of change and it proved very difficult to obtain extensive data on any particular measure of radar observer proficiency. Two main types of measures of proficiency were finally obtained for various groups of radar observers in the Continental Air Forces. These were ratings by instructors and radar bombing average circular error scores. A few additional measures were also explored with smaller number of crews.

Ratings by Instructors

Mission Ratings

Three samples of instructors ratings were obtained in the Fourth Air Force on radar observers in training at the Lead Crew School, Muroc Army Air Field, Muroc, Calif. The rating scale used was constructed by radar instructor personnel at the school, all of whom were combat returnees. The scale required instructors to rate the students on each of six items using a five-point scale. Weights were assigned to each of the items so as to result in a maximum score of 100 for all items. This score from all the items is referred to in the following discussion as the mission average rating. In addition to the scores on the items, the instructors also made an over-all appraisal of the radar observer, using a five-point scale. The reliabilities of the over-all and the mission average ratings are presented in table 9.1. Separate coefficients were computed for each sample between the ratings for odd and even missions. Only ratings by different instructors were included in these comparisons. These were corrected by the Spearman-Brown formula. The weighted average of the coefficients for all samples was then obtained.

TABLE 9.1.—*Reliability coefficients of average and over-all mission ratings of radar observer*

LEAD CREW SCHOOL, MUROC AAF, FOURTH AIR FORCE

Class	N	Average rating r_{II}	Over-all rating r_{II}
June	23	0.95	0.86
July	38	.40	.73
August	24	.68	.67
Weighted average	85	.63	.75

Two rating scales were constructed by research personnel in the Second Air Force to be used in evaluating the performance of radar observers. Both scales are described in some detail in the chapter on Lead Crews, Chapter 11. In Descriptive Rating Scale A, used in the 16th Wing, instructors were required to rate the radar observer trainees on a five-point scale on each of five items: scope interpretation, adjustment of equipment, maintenance of records, and lead crew potentialities. In Scale B, used in the 17th Wing, instructors were called upon to rate trainees in a similar manner upon these five items: lead crew potentialities, use of equipment, interest in duties, ability to work with others, and relative position of the trainee in his group. Distributions of ratings on the Scales are shown in table 9.2. Reliability coefficients for the two scales are shown in table 9.3. These were obtained by comparing the sum of the ratings given by one instructor with the sum of the ratings given by another instructor. In the computations of these coefficients, whenever a rating of a trainee on one item was omitted by a rater, the median rating given by that rater to that trainee on the other items was used for the omitted item.

TABLE 9.2.—*Distributions of radar observer ratings*

SECOND AIR FORCE

Scale A Rating	N	*Scale B* Rating	N
4	6	4–5	6
5	9	6–7	18
6	17	8–9	17
7	20	10–11	30
8	35	12–13	37
9	37	14–15	54
10	27	16–17	23
11	41	18–19	15
12	71	20–21	3
13	44	22–23	4
14	9	24–25	2
15	2		
16	1		
17	0		
18	1	Total	209
Total	320	Mean	12.95
Mean	10.18	S.D.	4.02
S.D.	2.54	Number of items	5
Number of items	4		

TABLE 9.3.—*Reliability coefficients of radar observer ratings*

SECOND AIR FORCE

Station	N	First rater, 1		Second rater, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Scale A:						
Alamogordo 3	23	8.30	2.73	9.83	3.06	0.86
Alamogordo 6	23	9.01	2.11	9.65	2.16	.48
Biggs	41	11.46	1.08	11.44	1.68	.76
Clovis	36	12.05	1.75	12.03	1.85	.42
Tucson	37	7.97	1.24	8.41	2.44	.23
Combined (average by Fisher's z)	160	9.98	2.46	10.38	2.62	.59
Scale B:						
Great Bend	28	2.53	0.64	2.64	0.83	0.52
Walker	32	2.64	.82	2.32	.91	.78
Albuquerque	19	2.90	.90	2.44	.58	.22
Combined (average by Fisher's z)	79	2.66	.79	2.46	.84	.59

Radar Average Circular Error

A detailed description of radar average circular error scores is presented in Chapter 7 in discussions of various bombing scores. Reliabilities for three radar bombing scores used in the Fourth Air Force are shown in table 9.4. These coefficients represent the correlation between average scores for odd and even missions corrected by the Spearman-Brown formula for each of several classes separately and for the total group. In the Second Air Force reliability coefficients were obtained by computing separately the corresponding T-scores for each type of radar bombing score: Photo Score (Victorville), Photo Score (Nadir) and Radar Score (Automatic). These T-scores were then combined into average over-all odd and even mission

radar CE T-Scores. Cases which did not have at least four scores in one type were discarded. Since T-scores were determined from distributions for each type by station, all cases from different stations were also combined directly. Reliability coefficients obtained in this way are shown in table 9.5.

Table 9.4.—Reliabilities of radar bombing scores on odd and even missions

MUROC AAF—FOURTH AIR FORCE

Measure	June		July		August		Total	
	N	r_{II}	N	r_{II}	N	r_{II}	N	r_{II}
Radar ACE total	23	0.61	38	0.76	41	0.68	102	0.[illegible]
Radar ACE normal aiming	23	.65	38	.70	41	.45	102	.50
Per cent gross errors	23	.20	38	.65	41	.61	102	.53

Table 9.5.—Reliability of over-all radar CE T-scores odd missions versus even missions

16th AND 17th WINGS—SECOND AIR FORCE

Station	N	Odd missions, 1		Even missions, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Alamogordo	20	51.90	6.62	49.40	6.02	0.55
Albuquerque	23	48.10	4.78	50.02	6.16	.36
Clovis	56	49.69	4.67	49.28	3.78	.28
El Paso	25	49.06	5.48	50.14	7.54	.28
Great Bend	31	51.76	5.72	48.98	5.80	.52
Tucson	38	49.72	5.56	49.14	6.18	.30
Walker	46	49.16	4.92	49.68	5.35	.54
Combined (average by Fisher's z)	219	49.84	5.48	49.48	5.80	.44
Coefficient of reliability (corrected by Spearman-Brown formula)						.61

Ground School Grades

Ground school grades in terms of "upper" and "lower" half of the class were obtained for each member in one class at each of four combat crew training stations (VH) in the Second Air Force. It was not possible to determine the reliability of these data.

Number of Radar Bombing Missions Completed

For one class at each of seven bases the number of radar bombing missions accomplished by crews within a specific period of time were determined. It was not possible to compute the reliability of this criterion of proficiency.

General Evaluation

Relations Between Criteria

The intercorrelations of three criteria available for radar observers in the Fourth Air Force are shown in table 9.6. The relationship of rating scale scores to other criteria available for radar observers are shown in tables 9.7, 9.8 and 9.9.

Evaluation

1. Taken as a group, the various rating scale scores were of sufficient reliability to be useful for validation purposes. The distribution of such

scores was good. It is recognized that the samples available for any one scale were small. However, from sample to sample the coefficients were fairly consistent. Descriptive ratings such as were used here are naturally open to the usual subjective sources of error. In this connection, it is likely that the mission ratings were less subject to such errors than over-all de-

TABLE 9.6.—*Intercorrelations of radar bombing criteria*

JUNE, JULY AND AUGUST CLASSES MUROC AAF—FOURTH AIR FORCE

	1	2	3	4	5	N
1. ACE total	......	0.74	0.87	0.19	0.20	102
2. ACE normal aiming	......	......	.45	.11	.09	102
3. Per cent gross error	......	......	......	.17	.17	102
4. Average rating	......	......	......	......	.87	99
5. Over-all rating	......	......	......	......	......	99

TABLE 9.7.—*Correlations between rating scale scores and ground school grades*

16th WING—SECOND AIR FORCE

Station	N_t	P_e	M_e	M_j	SD_t	r
Alamogordo 3	23	0.52	17.68	18.50	5.52	0.06
Alamogordo 6	23	.52	17.84	19.14	3.62	.23
Biggs Field	40	.52	23.50	24.54	2.80	.23
Clovis	36	.53	23.50	24.80	3.00	.27
Tucson	37	.51	16.02	16.60	3.16	.11
Combined (average by Fisher's z)	159	.52	20.12	21.06	4.75	*.19

TABLE 9.8.—*Correlations between total rating scale scores and radar terminal ACE T-scores*

17th WING—SECOND AIR FORCE

Station	N	Rating Scores, 1		Circular Errors, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Great Bend	31	25.82	7.52	50.11	5.32	−0.02
Walker	34	26.46	7.72	49.85	4.14	.11
Albuquerque	21	27.10	6.40	50.02	4.46	−.22
Combined (average by Fisher's z)	86	26.38	7.36	49.99	4.66	−.02

TABLE 9.9.—*Correlations of number of missions completed in radar/camera bombing training and total rating scale scores*

16th AND 17th WINGS—SECOND AIR FORCE

Station	Rating scale	N	Missions Completed, 1		Rating scores, 2		r_{12}
			Mean	S. D.	Mean	S. D.	
Alamogordo	A	45	2.93	2.75	9.15	2.43	0.02
Clovis	A	36	13.66	6.06	12.05	1.28	.08
El Paso	A	41	4.98	4.78	11.44	1.31	.29
Tucson	A	37	11.68	5.08	8.18	1.50	.15
Combined	A	159	7.92	6.51	10.17	2.33	.14
Great Bend	B	44	7.22	6.56	11.30	3.30	0.00
Walker	B	37	[illegible]	7.26	14.59	3.78	.05
Albuquerque	B	21	[illegible]	6.76	13.51	3.16	.27
Combined	B	102	11.08	7.61	12.95	3.76	.08
Scales A and B combined (average by Fisher's z)	....	261	......	......	......	......	0.11

scriptive ratings, since attention was directed upon performance rather than upon the individual. The average and over-all mission ratings in the Fourth Air Force were probably sufficiently highly correlated to warrant dropping one of these in the validation studies. Both scores also appeared to have the same relation to the various measures of bombing accuracy. However, since the additional work involved was not great, both measures were included in the few studies that were made.

2. Measures of bombing accuracy in both Second and Fourth Air Forces showed sufficient reliability to be useful in validation studies. Such scores have the great advantge of being little affected by subjective sources of error. Chief disadvantage of these scores were the frequent changes in method of determining bombing error and the presence of various types of clerical and computational errors in the available records. It should be noted that the radar bombing score reflected not only the proficiency of the radar observer but that of the pilot, navigator and bombardier as well. This point, frequently mentioned in connection with bombardier proficiency, is especially important in the case of radar observer proficiency because of the wide variation found among crews with regard to the role played by the radar specialists. In some crews, the radar member was the key person; in others, he played the role of a minor assistant. In some crews his duties were taken over completely by the navigator; in others, the radar observer performed major duties of both the navigator and bombardier. Thus, the adequacy of the bombing score as a criterion of proficiency for the radar observer may be seriously questioned, since no standard part of the bombing procedure may be assumed for all radar observers.

3. Ground school grades as criteria of proficiency for radar observers are subject to the same general criticisms that are made in the case of other aircrew positions. First, not all subjects are graded. Also many subjects that are graded have no obvious bearing on the trainee's task as a radar technician, being concerned with the fact that the technician is also a soldier. In no case was it possible to determine the reliability of either single grades or averages of all grades. In general, the prediction of ground school success is not likely to be a worthwhile enterprise unless eliminations from training are sometimes made on the basis of ground school performance. Such occurred very rarely in operational training.

4. The correlations between radar observer ratings and various measures of radar bombing accuracy are low. A possible reason for this is the fact that while the instructors attempted to rate the radar observer on his performance in the bombing operation, the actual bombing accuracy probably also depends upon the performance of several other crew members. Furthermore, the radar bombing scores were usually not known to the instructors who were making the ratings of radar observers. This is in marked contrast to the situation with respect to the bombardier where the average circular errors were known to and were frequently consulted by instructors rating bombardiers.

General Observations

The studies in this section are not strictly speaking validation projects in the sense that a specific selection procedure was being validated against success in a specialty for which it was designed. In the following studies, the bombardier, navigator and pilot stanines of radar observers are compared with a number of measures of proficiency. No evaluation is made of the validity of specific techniques developed by the Psychological Research Project (Radar) for the selection and classification of radar observers.

As was the case with other aircrew positions, research personnel found it necessary to spend considerable time in the evaluation of existing criteria of proficiency and the development of new measures. This analysis was considered necessary before validity coefficients could be appraised adequately. In the following discussion, validity coefficients significant at the 1 percent level (i.e. could occur by chance less than 1 time in 100) are identified by 2 asterisks and coefficients significant at the 5 percent level are marked with a single asterisk.

Stanine Validation

Rating Scales

Examination of table 9.10 suggests that, in general, the bombardier, navigator and pilot stanines have very low validity for predicting proficiency in radar observer training as measured by rating scale results. The 1 correlation coefficient significant at the 5 percent level, that between the bombardier stanine and rating scale score for the 16th Wing is not corroborated by results in the 17th Wing.

The correlation between the bombardier, navigator and pilot stanines of

TABLE 9.10.—*Correlations between stanines of radar observers and total rating scale scores*

SUM OF 2 RATERS, 16th AND 17th WINGS—SECOND AIR FORCE

Station	Stanine	N	Stanines, 1		Rating scores, 2		r_{12}	z[1]
			Mean	S. D.	Mean	S. D.		
Scale A:								
Alamogordo 3	N	22	7.50	1.23	18.40	5.40	0.31	......
Alamogordo 6	N	21	7.81	1.30	18.08	3.44	−.09	......
Biggs Field	N	41	6.98	1.35	23.04	2.80	.03	......
Clovis	N	34	7.41	1.38	24.08	3.08	.11	......
Tucson	N	37	6.50	1.32	16.26	3.10	.25	......
All Stations	N	155	7.23	1.34	20.18	4.36	.15	0.13
Do	B	155	7.22	1.36	20.18	4.36	.08	.16
Do	P+	155	6.00	1.52	20.18	4.36	−.09	.09
Scale B:								
Albuquerque	N	26	7.58	1.31	28.42	6.12	−.13	......
Walker	N	44	7.64	1.86	26.58	7.84	−.17	......
Great Bend	N	41	8.34	[illegible].12	27.10	7.32	−.02	......
All Stations	N	111	7.85	[illegible]6	27.22	7.32	−.11	−.11
Do	B	111	7.47	[illegible]	27.22	7.32	.02	.05
Do	P	11[illegible]	6.40	1.[illegible]	27.22	7.32	.11	.09

[1] Coefficients were combined for the stations by converting into "z" equivalents and also by combining all data directly.

radar observers at Muroc AAF and ratings by instructors are given in table 9.11. None of the coefficients is significantly different from zero.

TABLE 9.11.—*Correlations between ratings by instructors and stanines of radar observers*

66 RADAR OBSERVERS—MUROC AAF—FOURTH AIR FORCE

Type of rating	Bombardier stanine			Navigator stanine			Pilot stanine		
	Mean	S. D.	r	Mean	S. D.	r	Mean	S. D.	r
Average rating	7.65	1.51	0.03	8.29	1.05	0.04	6.26	1.87	−0.06
Over-all rating	7.65	1.51	.01	8.29	1.05	.04	6.26	1.87	.06

Average Circular Error

Tables 9.12 and 9.13 indicate that stanine validities are also consistently low when Radar Average Circular Error Scores are used as the criterion.

TABLE 9.12.—*Correlations between bombardier and navigator stanine of the radar observer and radar average circular error (T-scores)*

16th AND 17th WINGS—SECOND AIR FORCE

Navigator stanine

Station	N	Stanine, 1		Radar T-scores, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Alamogordo	26	[illegible]	1.31	50.14	5.66	−0.23
Albuquerque	17	7.41	1.42	49.02	4.88	.13
Clovis	33	7.42	1.39	49.83	3.51	−.18
El Paso	27	7.22	1.42	49.84	5.07	.20
Great Bend	29	8.14	1.36	49.64	5.18	−.10
Tucson	38	6.63	1.40	49.44	5.42	.06
Walker	29	7.62	1.71	50.08	3.67	.08
Combined (average by Fisher's z)	199	7.40	1.48	49.74	4.82	−.02

Bombardier stanine

Station	N	Stanine, 1		Radar T-scores, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Alamogordo	26	7.62	1.25	50.14	5.66	−0.03
Albuquerque	17	7.00	1.42	49.02	4.88	−.03
Clovis	33	7.21	1.32	[illegible]	3.51	.01
El Paso	27	6.78	1.13	49.84	5.07	.18
Great Bend	29	7.55	1.28	49.64	5.18	[illegible]
Tucson	38	7.13	1.44	49.44	5.42	−.16
Walker	29	7.00	1.58	50.08	[illegible]	.07
Combined (average by Fisher's z)	199	7.19	1.38	49.74	[illegible]	−.03

TABLE 9.13.—*Correlations between radar bombing scores and stanines of radar observers*

69 RADAR OBSERVERS—MUROC AAF—FOURTH AIR FORCE

Type of bombing score	Bombardier stanine			Navigator stanine			Pilot stanine		
	Mean	S. D.	r	Mean	S. D.	r	Mean	S. D.	r
Total average circular error	7.61	1.52	−0.01	8.26	1.13	−0.08	6.25	1.84	0.06
Average circular error normal aiming	7.61	1.52	.00	8.26	1.13	−.01	6.25	1.84	.04
Percent occurrence of gross errors	7.61	1.52	−.02	8.26	1.13	[illegible]	6.25	1.84	.06

Ground School Grades

In the single sample for which data were available, neither the bombardier nor navigator stanines were significantly correlated with ground school grades. The data are given in table 9.14. This result is in contrast

to that obtained in most of the other aircrew specialties where the navigator stanine usually had some correlation with ground school performance.

TABLE 9.14.—*Correlations between bombardier and navigator stanines and ground school grades of radar observers*

16th WING—SECOND AIR FORCE

Stanine	N_t	M_s	M_t	P	$S.D._s$	r_{bb}
N	154	7.22	7.25	0.53	1.45	−0.01
B	154	7.22	7.04	.54	1.44	.06

Number of Radar Bombing Missions Completed

The results presented in table 9.15 suggest that bombardier and navigator stanines are negatively related, (significant at the five percent level) to radar observer proficiency as measured by number of radar bombing missions completed within a specific period. This result is difficult to interpret since as discussed above, it is rather doubtful that the radar observer influences this measure to any great extent either favorably or unfavorably. However, analysis of the data in the table indicates that one class, Walker, with a high negative coefficient was the major source of these results. This coefficient is markedly different than those obtained with other samples, none of which provided coefficients significantly different from zero. It is possible that some spurious unknown factor was operating to produce these results, and that the true correlation between stanines of radar observer and number of missions completed is zero.

TABLE 9.15.—*Correlations between number of radar missions completed and stanines of radar observer*

SECOND AIR FORCE

Station	N	Missions, *1*		Bombardier stanine, *2*			Navigator stanine, *3*		
		Mean	S. D.	Mean	S. D.	r_{12}[1]	Mean	S. D.	r_{13}[1]
Alamogordo	44	2.86	2.69	7.59	1.34	.12	7.70	1.34	−0.06
Albuquerque	18	12.06	6.48	7.06	1.39	−.32	7.44	1.38	−.09
Clovis	33	14.08	5.06	7.21	1.32	−.10	7.42	1.59	−.21
El Paso	41	4.98	4.78	6.71	1.28	.01	7.02	1.41	.10
Great Bend	48	6.42	6.56	7.71	1.30	−.25	8.17	1.30	−.12
Tucson	42	11.16	5.44	7.17	1.38	−.11	6.69	1.42	−.07
Walker	32	14.26	7.18	7.13	1.38	**−.47	7.81	1.65	**−.55
Combined (average by Fisher's z)	258	8.70	6.96	7.26	1.37	*−.14	7.47	1.49	*−.15

[1] Positive sign of r's indicates a positive association between high stanine and high number of missions.

Evaluation

Stanines of radar observers show little relationship to the measures of proficiency discussed in this chapter. It is suggested that the inadequacy of the criteria was a major factor in this result. With the possible exception of rating scale scores (which are, in turn, subject to limitations of their own) no criteria were available which were dependent solely or even primarily upon radar observer proficiency. Further research in the development of measures of proficiency for radar observers is clearly indicated as

well as a better standardization of his duties. This lack of standardization reduces the confidence one would ordinarly have in the adequacy of radar bombing scores as criteria of radar observer proficiency. In connection with the low validities reported, it is worth repeating that the stanines on which these studies were based were not developed to predict the proficiency in radar observers.

SUMMARY

Research on radar observers was largely limited to examination of existing criteria and the development of ratings of proficiency in this specialty. Among the existing criteria only radar bombing accuracy scores were found suitable. Over-all descriptive ratings were tried out in the Second Air Force and mission ratings in the Fourth Air Force. None of the proficiency criteria that were studied were considered suitable as measures of radar observer performance. Stanines of radar observers showed little or no correlation with any of the available criteria.

CHAPTER TEN

Flexible Gunnery

ANALYSIS OF DUTIES AND SPECIFICATIONS

Job Analysis

No formal job analysis of flexible gunnery was made in the Continental Air Forces. However, certain general observations and analyses of details of operational training made in the Second Air Force are pertinent to the data which were obtained. In this connection it should be pointed out first of all that each gunner had duties in addition to the defense of the aircraft against attacks by hostile planes. Since the nature of these duties, as well as the nature of the gunners' task itself, were rather different in heavy and very heavy bombardment operations, a brief description of duties in the two types of aircraft is given in the following paragraphs.

B–17 and B–24 aircraft normally had six gun positions: nose, tail, upper and lower turrets and two waist guns. The upper and lower turrets utilized computing sights while the rest of the guns were hand sighted. Because the task involved was somewhat different, attempts were made to give all gunners experience in the operation of both types of sights. Until the fall of 1944 the usual heavy bombardment crew included six gunners including a radio operator, an aerial engineer and three aerial or career gunners. Since the bombardier manned the nose position, five of the gunners manned the five remaining guns and one usually served as a relief for the rest. In the fall of 1944, one person was dropped from the heavy bombardment crews. In half of the crews this was the bombardier, his place in the nose being taken by the armorer gunner. In the other half of the crews, one career gunner was eliminated. Two of the regularly assigned gunners, the aerial engineer and the radio operator had duties other than gunnery that were probably much more important. These positions are discussed in more detail in Chapter 8 and Chapter 11.

The normal B–29 crew included 5 gunners. These manned the tail, two waist or blister guns and the top sighting position. The radar gunner normally served as a relief for the rest. All gun positions had computing sights. However, the top sighting position represented a task rather different from single turret operation, since it served ordinarily for the remote control turret operation. The RCT gunner occupied this position and had been given special training for this job. The repair, inspection and operation of remote control turret equipment was his main responsibility. In addition to their gunnery duties, the two waist or blister gunners routinely served as scanners to observe the functioning of engines, control surfaces and landing

gear which were not visible to other crew members. While it did not involve any high degree of skill, this function was of vital importance and received great emphasis in training. In contrast to the heavy bombardment crew, the radio operator and flight engineer on B–29 aircraft had no gunnery responsibility. As with heavy bombardment crews, the nose position was manned by the bombardier. Certain features were common to both heavy and very heavy bombardment crews. All gunners had to be thoroughly familiar with their equipment and with any auxiliary equipment near their stations. They had to be able to detect and correct malfunctions of their guns and, if possible, make minor repairs in flight. Their job was naturally more difficult in B–29 aircraft since the equipment was more complicated. In the B–29 there was in addition a greater problem of coordination between gunners in the use of remote control firing equipment. Finally, the tail gunner had certain other responsibilities. In all formation flying, the tail gunner provided rear vision for airplane commander and crew. This aspect of his duty was especially important when his plane was in the lead position. In B–29 aircraft the tail gunner was also responsible for starting, operating and turning off the auxiliary electrical power plant or "put-put."

Gunners received three types of training in operational training: ground school courses, practice on synthetic trainers and firing at fighter planes in simulated interceptor attacks. The ground school courses included material on theory of ballistics, details of guns and equipment and their operation, and theories of formation defense and position firing. The type of trainer in use varied, but with installation in the fall of 1944 of E-14 or Jam Handy trainers at all stations, these became standard. Training on gunnery missions in the air included practice in operating all types of guns against a variety of attacks by fighter planes. Gun cameras mounted on guns or sights were used to evaluate the performance of the gunners. Sometimes blank ammunition was used to simulate combat conditions more closely. Minimum requirements with this type of training were the exposure of at least 50 feet of gun camera film on each of 4 missions with a minimum of 1 fighter attack recorded on each 25 feet of film.

Job Specifications

Except for the last year of the war, flexible gunnery received considerably less attention in operational training than did the other specialties of the combat crew. It was difficult, therefore, to obtain clear cut evidence as to what were the most important qualities for success in this task. Observations of the operation of central fire control equipment and of computing sights led to the conclusion that a high degree of skill in eye–hand coordination was definitely needed. The coordination required was certainly as complex as that required of the bombardier, for example. Another quality that was undoubtedly important for the group of gunners as a whole was teamwork. Coordination between individual gunners in defense against attacking aircraft was just as important as was coordination be-

tween other crew members in radar bombing and in successful operation of the aircraft on long distance missions. Alertness was particularly important for the blister gunners that acted as scanners. Their job was monotonous and tended too often to be neglected until after an emergency had already arisen.

CRITERIA OF PROFICIENCY

Ground School Grades

As with the other aircrew specialties, the management of ground school courses varied greatly from station to station. Subject matter and hours of training were fairly well standardized. Types of examinations and degree of subjectivity of grades given were the chief variants. The reliability of the average ground school grades of B-17 gunners at one station is shown in table 10.1. Also shown in the table are intercorrelations between scores on three examinations for students at a gunnery instructors school at Pueblo AAB in the Second Air Force. The reliability of one of these, the remote control turret examination, was determined by the comparison between scores on the odd and even items of the test. The mean score on the test was 83.01, the standard deviation was 7.40 and the reliability of the total test 0.74, based on data from 198 students. Ground school grades were obtained for 4 B-29 CCTS classes for whom ratings on Scale A were obtained. No reliability could be determined since final grade only was obtained.

TABLE 10.1.—*Reliability coefficients of average ground school grades—all courses*

B-17 GUNNERS—SECOND AIR FORCE

A = Aerial engineer and radio operator. B = Armorer gunner and career gunners.

Type of gunner	Class	Number of cases	Grade phase II, 1		Grade phase III, 2		r_{12}
			Mean	S. D.	Mean	S. D.	
A	12-10	112	76.9	5.72	82.7	4.14	0.43
B	12-10	196	77.9	5.56	85.9	5.02	.41
A	4-1	111	80.8	5.92	82.2	5.20	.69
B	4-1	190	83.9	3.48	84.0	3.20	.59

Intercorrelations between test scores

GUNNERY INSTRUCTORS SCHOOL—172 B-29 GUNNERS

Examination	Mean	S. D.	Code	A	B	C
A. Remote control turret	83.3	7.37	A	.....	0.34	0.26
B. Crew coordination	82.4	9.14	B	.....	.....	.31
C. Formation defense	82.8	12.50	C	.....	.....	.....

Ratings by Instructors

Although the importance of such duties was everywhere recognized, there were no records available which gave any evidence of the degree of proficiency of gunners in carrying out duties other than gunnery. Subjective ratings by instructors were thus the only means of obtaining estimates of

this proficiency. In connection with a study of lead crew proficiency two rating scales were constructed in the Second Air Force and administered to a number of B-29 crews in that Command. In each case two relatively independent ratings were obtained on each crew rated. Each scale contained items referring to the gunners of the crews being rated. In both scales the gunners were referred to as a group rather than individually. Ratings were handled in this fashion because it was believed that individual gunners were not well enough known to be rated individually. However, at Biggs Field instructors using Scale A offered to rate each gunner individually on each of the two items referring to gunners. Two independent ratings were thus obtained at Biggs Field for 164 gunners of 41 crews. In table 10.2 are given the correlations between these ratings. Since at this station, gunners were regarded as interchangeable between positions, separate coefficients were not computed for different positions. It is apparent from the results shown in the table that the original hypothesis that individual gunners were not well known to instructors is supported. The data in table 10.3 suggest, however, that instructors agreed moderately well in their ratings of the gunners as a group.

TABLE 10.2.—*Correlations between independent ratings of gunners*

BIGGS FIELD—41 CREWS—164 GUNNERS

	First rater, *1*		Second rater, *2*		r_{12}
	Mean	S. D.	Mean	S. D.	
Item 1	2.22	0.66	2.35	0.60	−0.04
Item 2	2.23	.65	2.43	.59	.00
Items 1 and 2	4.45	1.33	4.80	.85	−.06

TABLE 10.3.—*Reliability coefficients of ratings of gunners with gunners of a crew rated as a group*

B-29 CREWS—SECOND AIR FORCE

Station	N	First rater, *1*		Second rater, *2*		r_{12}
		Mean	S. D.	Mean	S. D.	
Scale A:						
Alamogordo 3	23	5.35	1.16	6.48	1.09	0.66
Alamogordo 6	23	5.03	1.12	5.00	.83	.42
Biggs	41	4.66	.93	5.07	.71	.29
Clovis	36	6.17	1.70	6.83	1.26	.30
Tucson	37	5.57	1.32	5.70	1.78	.61
Combined (average by Fisher's z)	160	5.36	1.40	5.80	1.44	.45
Scale B:						
Walker	17	2.24	.36	3.30	.53	0.10
Albuquerque	49	2.44	.96	2.47	.92	.86
Pyote	27	2.92	.40	2.72	.42	.30
Combined (average by Fisher's z)	93	2.62	.73	2.69	.83	.66

The biserial coefficients presented in table 10.4 between ratings and ground school grades (with students divided into upper and lower groups) indicate that, for five CCTS (VH) classes in the Second Air Force, grades and ratings were not correlated.

Trainer Scores

The type of gunnery trainers used and trainer requirements changed frequently during the period of psychological research in the Continental Air Forces. Since the E-14 or Jam Handy trainer was finally selected as standard most of the data obtained are from this trainer.

The task of the gunner with this trainer was to follow with his sights an image of a plane projected upon the trainer screen. Accuracy of performance was determined by an instructor or observer who watched the gunner in action. To provide a more accurate estimate of the amount and type of error, the instructor had a switch or button with which he could project a reticle upon the screen at the aiming point of the gunner. By visual comparison with the position of the plane image, it was possible to estimate the degree and direction of gunner errors. The apparatus and

TABLE 10.4.—*Correlations between ratings (Scale A) and ground school grades*

Station	N	P_e	M_e	M_t	SD_t	r_{te}
Alamogordo 3	23	0.35	12.50	11.90	2.16	−0.16
Alamogordo 6	23	.43	10.20	10.62	1.31	.20
Biggs Field	41	.88	9.46	9.60	1.33	.08
Clovis	36	.47	12.68	13.40	2.44	.19
Tucson	37	.54	11.20	11.56	2.88	.07
Combined (average by Fisher's z)	160	.57	10.79	11.82	2.48	.07

method of determining error were different for B–17 and B–29 training, since the type of sighting used was different. The Jam Handy equipment paralleled hand sighting in B–17 training and, in B–29 training, it was similar to the computing sight. The technical details would require a lengthy explanation if presented here and are regarded as beyond the scope of this report.

In B–17 and B–24 training each gunner was given at least three formal tests or "phase checks" on the E-14 trainer during operational training. A test or phase check consisted of a 3- to 4-minute period on the trainer during which an instructor observed the gunner's performance. During each period the instructor made a total of 10 observations on the accuracy of sighting. The score for each observation was given on a 10-point scale, each point being equal to 1 "rad" of error. Ten was to be a perfect score, and zero the lowest possible score which represented 10 or more "rads" of error. The scores for the 10 observations added together formed the total test or phase check score. Each gunner was given a test or phase check as near the beginning of a phase of training as possible.

Prior to December 1944 all gunners were supposed to have the same amount of practice on the Jam Handy. It proved to be very difficult to schedule the required practice sessions for all gunners. Both instructors and equipment were overloaded. Hence, in December 1944, it was decided to set 95 as the minimum test score, and to excuse from further practice in that phase of training all gunners equalling or exceeding that score on the

test for the phase. A second objective was to improve the motivation of the gunners in their practice on the trainer. In this connection, it is interesting to note the increase in score on the Jam Handy test for Phase I following this change in procedure. The pertinent data are given in table 10.5.

TABLE 10.5.—*Scores on test 1 of the Jam Handy trainer*

BIGGS FIELD—B-17 GUNNERS—SECOND AIR FORCE

	No. cases	Mean score	S. D.
Before December 1944	162	84.34	7.72
After December 1944	159	94.64	2.79

The reliability of Jam Handy test scores obtained as described above is illustrated in data obtained from two B–17 CCTS. The correlation between test 1 and test 2 for Biggs Field and the average intercorrelation for from two to four tests for Ardmore are given in table 10.6. It is obvious that these scores have little or no reliability.

TABLE 10.6.—*Reliability coefficients of B-17 Jam Handy test scores*

SECOND AIR FORCE

Station	Number of cases	Score first test, 1		Score second test, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Biggs	183	91.3	29.2	94.8	29.2	−0.11
Ardmore	215	......	......	......	......	[1].06

[1] Average of intercorrelations, 2–4 tests.

Not only did the B–29 crew training on the Jam Handy differ from that for B–17 and B–24 crews in apparatus and method of sighting, but the method of scoring also differed. Little attempt was made to secure a numerical estimate of the degree of error in aiming. Subjective ratings were made by the instructor on three aspects of performance: framing, tracking and burst control. Ten observations were made on a gunner in each test or phase check. The ratings for the three aspects of performance were added together for the ten observations and the score was then the percentage which this value was of the maximum possible total score. Students were not excused from training on the basis of a minimum score and the number of tests given varied from student to student. In a group of B–29 gunners at Alamogordo the number of tests ranged from two to nine. In table 10.7 are given the correlations between scores achieved on odd and even tests. Although the reliability is very low, there is some evidence that with a larger sample of performance, a higher reliability might be obtained. The fact that the number of tests varied may have made these values spuriously high. However, no more detailed analysis was believed warranted by the available data.

Data on the Jam Handy test scores were obtained from one CCTS (H) class where the standard scoring procedure developed at the Central School for Flexible Gunnery at Laredo AAF was used. This procedure was in process of being set up throughout the Second Air Force at the close of

hostilities. The main points of difference from current practices were a better definition of errors and the inclusion in each test of 20 instead of 10 observations. Also special standardized test films were used. In table 10.8 are given the reliabilities computed within tests and computed between tests for each of the gunners of the crews. As might be expected the within

TABLE 10.7.—*Reliability coefficients of B-29 Jam Handy test scores*

ALAMOGORDO—B-29 CREWS—SECOND AIR FORCE

	Number of cases	Odd test score		Even test score		r_{12}
		Mean	S. D.	Mean	S. D.	
2 or more tests	470	79.3	5.76	81.9	6.02	0.06
4 or more tests	202	80.8	4.78	81.5	5.68	.41

test reliabilities are high while the correlation between test scores is low. However, it should be noted that these are still higher than those obtained under more common current practice where the procedures were less standardized.

TABLE 10.8.—*Reliability coefficients of Jam Handy test scores*

LAREDO ARMY AIR FIELD—SCORING PROCEDURE

Within-test reliability, three tests each gunner

Gunner	Number of cases	Odd scores		Even scores		r_{12}
		Mean	S. D.	Mean	S. D.	
Aerial engineer	55	86.8	4.74	86.8	3.93	0.71
Radio operator	56	87.2	4.36	87.0	4.29	.82
Waist gunner	28	84.6	4.69	85.5	5.10	.37
Ball gunner	55	87.0	4.23	87.3	4.04	.48
Armorer gunner	56	88.6	4.03	88.3	4.08	.73
Tail gunner	56	86.0	5.78	84.8	6.14	.86
All gunners	306	......	......	......	......	[1].72

[1] Weighted "z" average of all positions.

Between-test reliability, "z" average of intercorrelations of three tests

Gunner	Number of cases	Test 1		Test 2		Test 3		r (av.)
		Mean	S. D.	Mean	S. D.	Mean	S. D.	
Aerial engineer	52	163	15.0	176	8.3	181	5.1	0.23
Radio operator	50	165	14.0	175	9.5	181	5.9	.05
Waist gunner	22	153	17.0	170	7.1	181	5.9	.08
Ball gunner	47	169	11.6	172	8.9	179	5.8	.22
Armorer gunner	51	171	10.3	177	9.1	181	7.6	.15
Tail gunner	51	160	16.0	172	9.9	179	16.2	.44
All gunners	273	......	......	......	......	......	......	.21

Gun Camera Scores

As with the scoring of trainer performance, scoring of gun camera missions varied greatly from station to station in the Second Air Force. A great deal of difficulty was present at first in securing adequate camera equipment. Different cameras differed somewhat in their operation. Frequently camera malfunctions prevented the obtaining of assessable films. One common practice made most of the available data of doubtful value. Because of the difficulty of obtaining sufficient missions to meet minimum

requirements non-assessable film was frequently credited toward meeting these requirements. Since no distinction was made between film that was nonassessable because of camera or other unavoidable malfunctions and film that was nonassessable because the attacking aircraft did not appear in it, the assessed film represented a rather selected sample of gunner performance.

The scoring procedures used were very different for hand aimed guns and guns with computing sights. Scoring of film obtained with hand aimed guns offered tremendous difficulties and it is doubtful whether the scores bore any relation to probable accuracy of shooting. The actual score given on a mission was a sort of percent hits score, based on a subjective judgment by an observer viewing the film projected on a screen. The person scoring the film picked out what he thought were typical frames for each attack by a fighter aircraft. For each typical frame the scorer decided whether the reticle of the gun sight was correctly placed on the picture with respect to the attacking aircraft. If the sighting was judged correct, the frame was scored as a hit. If incorrect, it was scored as a miss. The judgment required of the scorer was exceedingly complex and, as regards true likelihood of hitting the attacking plane, impossible. Such factors as relative speeds and directions of movement of attacking and attacked planes were not represented in any way in the pictures obtained. About all a skilled scorer could do was to assume typical speed and direction of movement and decide whether the "lead" given in sighting was reasonable.

With computing sights a better assessment of performance was possible. In scoring this film a sample of the frames was also utilized. (The fact that the corrections in direction in which the guns actually pointed were automatic and the gunners' job was to set the sight directly on the attacking plane made scoring a much more simple process). This sample usually included a little over 5 percent of the frames registering the pursuit curve portion of the attack. The frames chosen for scoring were selected by the scorer who was instructed to select typical frames distributed about evenly through the total pursuit curve portion of the attack. Two types of errors were estimated for each frame scored. A tracking error was obtained by measuring on the photograph the distance from the center of the reticle of the sight to the center of the wing section of the fighter plane. A framing error was scored by measuring the difference between the width of the wing span of the attacking plane and the width of the reticle. The unit of measurement in each case was one-eighth "rad" or one-eighth of the radius of the first ring on the 35-mm. sight. After determination of the two types of error on each frame, there were several procedures in use to determine the score for the total mission. The gun camera scores obtained at the different stations were thus expressed in terms of various types of units, depending upon the particular procedure in use at those stations.

A detailed study was made of the gun camera records for one CCTS (H)

class at Sioux City. Training requirements were that each gunner expose at least 50 feet of film on each of 4 missions. In addition, the navigator and bombardier also exposed some film. The crews completed their gunnery requirements in an average of 6 missions per crew. Each crew exposed an average of 30 units of 50 feet of film (S.D. of 3). The exposed film represented a total of 3,350 attacks. The number of fighter attacks was actually much less than this, since a single attack on a formation of B–17 aircraft might be filmed by 15 or 20 gunners. As far as gunners were concerned the figure of 3,350 represented the total possible number of scores. Of these only 645 were scorable, so that the gun camera scores secured represented only 19 percent of the gunnery accomplished by this class of B–17 crews. In table 10.9 are given the reliabilities of the gun camera scores for this class. Within and between mission reliabilities are given for both hand sights and computing sights. In this study the hand

TABLE 10.9.—*Reliability coefficients of gun camera scores*

B–17 GUNNERS—SIOUX CITY AAB—SECOND AIR FORCE

Within mission reliability

	Number of cases	First attack, 1		Second attack, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Hand sighted guns	112	34.7	17.2	37.1	18.7	0.86
Computing sights:						
Framing	129	2.02	.86	1.97	.86	.60
Tracking	129	1.73	.97	1.68	1.10	.80
Total	129	1.79	.74	1.69	.71	.83

Between mission reliability

	Number of cases	First mission, 1		Second mission, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Hand sighted guns	31	35.8	18.8	41.8	17.6	0.04
Computing sights	27	1.66	.59	1.38	.58	.46

BIGGS FIELD

	Number of cases	Odd missions, 1		Even missions, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Computing sights, 2 to 6 missions	210	32.2	4.92	32.5	5.04	0.18

Correlations between gun camera scores and ratings of film

	N_t	M_s	M_t	P	T	r_{bis}
Tracking consistency	204	33.8	31.6	0.33	3.82	0.36
Framing consistency	204	33.0	31.7	.47	3.82	.22
Self-inflicted damage	205	32.2	31.7	.57	4.06	.07

B–29 GUNNERS—PYOTE AAF

	Number of cases	First mission, 1		Second mission, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Two missions	151	32.2	4.92	32.5	5.04	0.18

sight scores were as described previously and the scores for the computing sights were the actual error scores obtained for the separate frames scored. No attempt was made here to use the routinely reported total gun camera score. As might be expected the within mission reliabilities were high and the between mission reliabilities were low. Reliability between missions was practically zero for the hand sights but was appreciable for the computing sights. The small numbers of cases make the latter result open to question.

Also given in the table is the correlation between scores on odd and even missions for 210 B–17 gunners at Biggs Field, each gunner having from 3 to 6 scored missions. In obtaining gun camera score for a mission, the average tracking and average framing error were first computed for all attacks and all frames scored from that mission. Then the equivalent number of hits per attack was read from a standard chart prepared for this purpose. For the same gunners there were available subjective ratings of consistency of tracking and framing and of presence or absence of self-inflicted damage. Correlations between these subjective ratings and average gun camera score are also given in table 10.9. Since the judgments were two-category ratings, all coefficients are biserial in form.

The reliability of gun camera scores for B–29 gunners at one Second Air Force Station is also shown in table 10.9. The procedure for obtaining total gun camera score at this station differed from that described for the B–17 data reported above. In addition to the determination of an average tracking and framing error as described earlier, a third error, percent of framing error, was computed by dividing the framing error by the target span. A fourth score, total error was really the final score for each frame and was obtained by multiplying the percent framing error by 50 (except that in waist position the factor was 100) and adding the tracking error. The total score for the mission was then 100 less the average of the scores from all frames scored.

Relation Between Trainer and Gun Camera Scores

In table 10.10 are given the correlations between gun camera scores on the Jam Handy and the E-8 trainers. The E-8 ("Spotlight") trainer had a spot of light moving over a screen in an erratic pattern and the task of the student was to follow the spot with the gun sight. When the sight was properly operated, pressing the burst control switch activated counters which recorded two scores; number of shots fired and number of shots fired while the sight was on the target. From these two were obtained the percent hits or total score. Only the tracking feature of gunnery was involved in operation of this trainer. From the data in the table, the E-8 trainer would seem to have a greater similarity of performance to aerial gunnery than the Jam Handy. The latter was selected for standard use at all stations primarily because E-8 equipment was difficult to standardize and keep in adjustment.

TABLE 10.10.—*Correlations between gun camera and gunnery trainer scores*

SECOND AIR FORCE

	Number of cases	Jam Handy, 3–4 scores, 1		Gun camera, 3–6 scores, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Biggs Field, B-17 crews	105	31.8	4.16	9.22	2.25	0.06

	Number of cases	E-8 trainer, 2 scores, 1		Gun camera, 2 scores, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Pyote AAF, B-29 crews	151	79.7	13.0	83.1	8.95	**0.26

SUMMARY

Since the primary missions of research personnel in the Continental Air Forces were related to the validation of selection and classification techniques used in the Training Command, little research was done in flexible gunnery, where such techniques were developed too late for use in validation studies. The research that was accomplished was necessitated by another mission assigned to research personnel, also late in the war, that of establishing procedures for the designation of potential lead crews. The studies made were primarily analyses of proficiency measures and synthetic trainer scores.

The results presented indicate that with the training methods in effect at the time, it was not possible for instructors to make reliable ratings of individual gunners. Ratings of the gunners of a crew as a group were more reliably made. Ground school grades were moderately reliable. The correlation between these two scores was negligible. The Jam Handy trainer, when procedures for its use were carefully controlled, provided scores of low but statistically significant reliability. Gun camera procedures were so varied and unstandardized that the resulting scores were of doubtful value. These scores had almost no reliability for hand-sighted guns; for computing sights their reliability was low but significant. The E-8 trainer provided scores which correlated more highly with gun camera scores than did those from the Jam Handy. The results are regarded as highly tentative in view of the low reliability of the variables concerned.

CHAPTER ELEVEN

Selection and Evaluation of Lead Crews

INTRODUCTION

The Tactical Importance of the Lead Crew

Although the advent of the atomic bomb may well change all tactical concepts in warfare, prior to its successful operation the lead crew occupied a peculiarly important place in bombing operations. Two quotations illustrate this very well. The first is a quotation from Manual 50-1, 20th Air Force, entitled, Lead Crew:

> "Experience in actual combat has demonstrated that skillful leadership is necessary if compact defensive formation is to be maintained and accurate, concentrated bombing of targets achieved. Experience has shown that bombing results have varied with the skill and experience of lead teams. . To place a high percentage of the bombs of an entire squadron or group within that circle, (within 500 feet of assigned MPI), is not a simple task, but one which involves a number of highly developed techniques. . . The burden of attaining this goal rests largely upon lead crews."

Major General Curtis LeMay, Commanding General of the XXI Bomber Command, states in the Foreword to the Lead Crew Manual of that Command:

> "The mission of this Command is to destroy enemy targets. Since our effectiveness in accomplishing this mission must be measured in terms of our *best,* rather than our *average,* capabilities, selected 'Lead Crews,' representing the best we have, will lead our formations. The success of our tactical operations, therefore, will depend to a high degree on Lead Crew proficiency. In no other military operation does so great a responsibility devolve on so small a group of specialists as in the case of Lead Crews. During the brief span of the bombing run, all of the human life, labor, and matériel that have been invested in placing the formation over the target are held 'in trust' by the Lead Crew, whose performance determines whether or not the investment pays off."

So much importance was attached to the part played by lead crews that special training was given crews chosen to serve in this capacity. This training was given at lead crew schools set up for this purpose in the 8th and 20th Air Forces. In other Air Forces it was a part of the tactical training given to organizations assigned to them. In a sense this training represents the most advanced step in the training of combat crews.

In the early part of the war, as doctrines were developed for the utilization of outstanding crews in lead position, it became more and more important to identify such crews early in their tours of combat duty. It was

obviously impossible to try out every crew in lead position. The whole procedure would work most effectively if some means were at hand to designate the most likely crews at the start of combat operations. These could then be watched closely and when they had the necessary experience could be tried out in lead position. It was only natural, therefore, that commanding officers receiving replacements in the theaters of operations should desire more and more information as to the proficiency of these crews. At first such information was sent sporadically and informally in inclosures to basic records sent overseas. Later the problem was recognized officially and plans were made to provide routinely the needed information.

Job Analysis

The duties of lead crews and their crew members included, of course, all of the regular duties of each specialty in any crew. However, some of these duties were much more important in a lead crew; and the thoroughness with which all duties were performed was also more important. In the paragraphs which follow some of the important aspects of the duties of lead crews are touched on very briefly. A much more detailed presentation can be obtained from Heavy and Very Hearvy Bombardment Training Manuals and from Manuals of Lead Crew Training.

Since the lead crew started all parts of and set the pace for the bombing operation for all of the crews, the first step in the preparation of lead crews for missions was obviously mastery by the crew members of every detail of the mission plan. Particularly important here were the assembly plan, the signals to be used, details of course decided upon to avoid flak and the directions for the attack upon the target. As with all crews, a thorough preflight inspection by the airplane commander of the crew and its equipment was essential.

On the mission the airplane commander had to make certain that the lead plane flew exactly according to mission S.O.P. and on exactly the briefed course. He saw to it that his own group maintained a close compact formation. He was responsible for having his crew monitor the VHF channel at all times. He checked frequently with the navigator as to progress on the course. At the IP he was required to turn the ship over to the bombardier well-trimmed and with the autopilot functioning properly. After the bombing run he had to so regulate the return flight that as many as possible of the stragglers could be protected; but he could not do this at the expense of the proper defensive formation.

The lead navigator had to make a particularly careful preflight check of all his instruments. He had to go over the entire mission with the crew, discussing every feature. On the flight he had to hit each control point on time. Not only did he have to gauge the path of his own plane with respect to the effects of wind, speed on turns, and other factors, but he had to take into account the flight characteristics of whole formations. In cooperation with the bombardier, (and radar observer) he had to pin point the position

of the plane frequently and give the appropriate information to the airplane commander from time to time. When approaching the IP the navigator alerted the bombardier and cooperated with him to make sure that the IP was identified early and certainly. He maintained not only the usual navigation logs but made records of weather, fighter opposition, antiaircraft fire and presence of friendly or hostile aircraft. At the actual bombing run he made a record of all possible details including the results of the bombing. He had to be prepared to take evasive action when passing points of resistance. Any changes in plans necessitated by delays or other circumstances had to be reported to the airplane commander for decision as to action.

The radar observer alternated the main coordination of his activities between navigator and bombardier. On the flight to and from the target he maintained a plot of the course as did the navigator. He assisted in directing the lead plane and the formation around known flak areas and the like. Much of his work was an independent check upon that of the navigator. On over-water flights, depending upon the presence of an undercast, the radar observer or the navigator reported to the other the moment of landfall and the ground speed at the time. Except to report and check with each other, the radar observer and navigator worked independently, thus assuring the most accurate possible navigation. As with other crew members a careful preflight check of the equipment was essential.

The bombardier like the navigator and radar observer checked all his equipment, including bombs, fuses, sight and bomb bay doors prior to takeoff. He also set his altimeter and checked his interphone position. He also briefed the crew regarding the bomb run itself. During the flight to the target the bombardier set up as much of the information as possible in his sight, checking such items as temperature and wind as he went along. He read drift from his sight to check on the drift obtained by the navigator He kept watch for friendly and enemy fighters and checked the crew on the oxygen equipment. He aided the pilot in setting up the autopilot. He checked and rechecked all data for accuracy. Prior to arrival at the IP he made trial runs, if possible, to check the set-up of computer and autopilot. He observed the terrain over enemy territory. Finally he cooperated with navigator and radar observer in locating the IP and then the target itself.

Depending upon the visibility the radar observer did or did not play a large part in the bomb run itself. If there was any difficulty in seeing the IP visually, the radar observer took the pilot around the IP on the turn toward the target. The bombardier meanwhile set up his sight at the direction of the radar observer as if he were bombing visually so that he could take over at any moment if the target came into view. If the target did not appear visually he set the sight into operation as if it had happened, using the signal of the radar observer and data supplied by him for its operation. The bombs were then dropped under direction of the latter.

The airplane commander, navigator, bombardier and radar observer

were frequently referred to in training directives and in conversations among training personnel as the "bombing team." However, the division of duties among the last three of these was, in actual practice, not as clearly defined as the above brief description might imply. Crew doctrines involving radar bombing were continually being developed and revised. Thus there was considerable variation from crew to crew as to which member of the bombing team performed a particular task. For example, in the B-29 aircraft the navigator had a radar scope and it was possible for him to make the turn and direct the bombing run, when the target was obscured. On many crews the navigator actually did so. Whether the bombardier, radar observer or navigator assumed the major duties on the bombing run appeared to be a matter worked out in part by the crew involved.

The radio operator contributed far more to the efficiency of combat crews in general and lead crews in particular than the status of radio operators would indicate. He was especially important in lead crews since all communications with the rest of the formation, with other formations and with appropriate ground stations passed through him. He was responsible for the proper functioning of all radio equipment and had to see that sets were tuned and operating properly at take-off. He checked spare fuses and tubes. The radio operator also usually inspected guns and other equipment. Where possible during the mission he attempted to code messages ahead of time so that no time would be lost in sending when the signal was given.

The final crew member of special importance for the lead crew was the tail-gunner. In many lead crews this was an officer—usually a copilot or a gunnery officer. Whether officer or enlisted man, he had to have a thorough knowledge of formation SOP and tactical doctrines as it was his job to report to the airplane commander the conduct of the formation behind the lead plane. His position was particularly important during assembly when he reported the positions of the various units in sight.

Job Specifications

The qualities necessary for success in lead crew position were given considerable study. Information regarding them was obtained from a number of sources. Aviation psychologists in each of the Continental Air Forces interviewed officers with combat experience, especially experience in lead crew position. In the Fourth Air Force instructors at the Lead Crew School at Muroc Army Air Field, Muroc, Calif. were interviewed. Additional information was obtained from training manuals, manuals for lead crew training, from study of intelligence reports and from published analyses of combat operations. An attempt was made especially in the interviews in the Second Air Force, to obtain traits and characteristics referring to the crews as a whole. Although most of the characteristics mentioned obviously referred to particular crew members, a few characteristics and traits referrable to the total crew were given considerable emphasis.

The instructors at the Lead Crew School at Muroc were in general agreement that the ideal lead crew was composed of men, all of whom were exceptional in their respective assignments. Expertness on the part of all personnel was not regarded as sufficient, however. Regarded as of equal importance was demonstration of an unusual amount of teamwork and cooperation on the part of individual crew members. A crew of independent, noncoöperating experts would not be regarded as a good lead crew. The typical manner in which, as a group, the instructors determined whether or not a crew was likely to be a lead crew, provides insight into their conception of the lead crew. They determined first whether each member of the crew possessed the fundamental skills of his aircrew postion to a marked extent. They then evaluated the amount and quality of the teamwork and coordination demonstrated by the crew. The relationships between pilot and flight engineer, pilot and navigator, and among the navigator, bombardier and radar observer were regarded as particularly important. The presence of an average or even a weak aircrew member would tend to be overlooked if the teamwork exhibited by the crew as a whole was extremely good.

Lead crew personnel interviewed in the Second Air Force agreed very well with the instructors at Muroc as to the most important characteristics for success as a lead crew. *Outstanding proficiency* in each specialty was regarded as of major importance. Also, as at Muroc, cooperation and teamwork were considered nearly as important. Certain additional characteristics were both frequently mentioned and given great weight in these interviews. The first of these was *combat experience*. It was felt that a good lead crew should have successfully met a wide range of combat conditions: heavy flak, fighter opposition, exhaustion of gas, high altitude, poor visibility, bad weather and equipment malfunctions. In a sense such a crew might be said to have proven itself and would thereby be better suited for lead position than an untried crew.

Another important characteristic was *eagerness and enthusiasm*. This was thought to involve a combination of strong desire to destroy the enemy and active enjoyment of the job of lead postition. The eager lead crew was thought of as one that flew extra missions, worked extra hours on trainers, studied briefing materials and learned each other's jobs. Related to this trait or characteristic were confidence, both in the airplane commander and in the crew as a crew, and pride in crew accomplishments.

A third quality considered important was *crew discipline*. In the well disciplined crew individual crew members were not unduly familiar with the airplane commander. They were always on time at formations and rarely had absences. They did not argue with each other over decisions made. Orders were obeyed without repetition. Crew members remained calm and worked smoothly and efficiently regardless of the situation. Interphone discipline was thought especially important. On the one hand there was no jabbering of useless material over the interphone and on the other

hand the airplane commander was kept informed of all important events and crew members checked frequently on each other.

Finally, *leadership*, especially of the airplane commander was considered important. A crew in which other crews had confidence was sometimes considered a better bet for lead crew than more able crews lacking such confidence. A good lead crew had few complaints from other crews in the formation after missions were over. Because of the prestige it gives, rank was mentioned as desirable in the lead crew.

The characteristics or qualities of individual crew members thought especially important for a lead crew are given in table 11.1. The table is a composite of material obtained in all three of the Continental Air Forces dealing with bombardment crews. No attempt is made to present the individual characteristics in order of importance since adequate information of this sort was not obtained in any Air Force. The one exception is proficiency in the individual specialty. This was clearly considered the most important single characteristic for each aircrew position in the lead crew. Because it applied to every crew position it is stated here and does not appear in the table under the individual specialties. It should, however, be understood that under each specialty this is the most important item.

TABLE 11.1 *Qualities important in lead crews*

Airplane commander:
- Formation flying:
  - Knowledge.
  - Foresight.
  - Smoothness of control.
- Experience in combat.
- Ability to plan ahead.
- Skill in instrument flying.
- Perceptual judgment.
- Interest in crew.
- Rank.
- Enthusiasm for the job.
- Calmness in emergencies.
- Leadership.

Copilot:
- Formation flying:
  - Foresight.
  - Smoothness of control.
- Instrument flying skill.
- Cooperation with airplane commander.

Navigator:
- Skill in celestial navigation.
- Skill in dead reckoning navigation
- Complete accuracy in pin-point navigation.
- Knowledge of formation flying.
- Foresight.
- Thoroughness in routine procedures.
- Good cooperation with crew.

Bombardier:
- Skill in dead reckoning navigation.
- Consistency of bombing.
- Foresight.
- Good target identification.
- Stability under stress.
- Enthusiasm for the job.
- Good cooperation with others.

Flight Engineer:
- Ability to meet changes in plans.
- Accuracy in routine computations.
- Skill in estimating.
- Ability to handle crew.

Radar observer:
- Skill in interpreting instruments.
- Ability to make fine adjustments of instruments.
- Proficiency in navigation.
- Good cooperation with crew.

Radio operator:
- Skill in sending and receiving.
- Skill in adjustment of equipment.

Gunners:
- Skill in judging distance and rate of movement.
- Thoroughness of preflight procedures.
- Cooperation with rest of crew.

Assembly of Potential Lead Crews

The need for providing overseas air forces with information about the proficiency of the crews being assigned there was officially recognized with the publication of AAF Letter 50-117, 7 June 1945. This letter went even further, however, and also specified procedures for the assignment of individuals to crews prior to operational training in an attempt to obtain the maximum possible number of outstanding crews. According to the provisions of this letter, personnel were assigned to crews at the AAF Combat Crew Processing and Distribution Center in the AAF Training Command on the basis of a lead crew aptitude score. This score was based on evaluation of aerial training and experience, on written examinations and on original aptitude scores. Individuals with the highest lead crew aptitude were assigned together on crews and such crews were tentatively designated potential lead crews. At the end of operational training the letter provided that the organization in which training was done would review the designations of the crews and either confirm or revise the lead crew designations for crews so designated prior to the training. This information was then to be sent overseas with the crews concerned. An appropriate form was provided for the transmission of the necessary information from the AAF Training Command through operational training to overseas commanders. The end of the war made the carrying out of these procedures unnecessary. For details of the provisions for crew assembly and the records to be used see AAF Letter 50-117, a copy of which is included in the appendix H.1.

Problems in Crew Evaluation

The concept of the team and emphasis on team spirit have been for years an important part of the social pattern in the United States. They find expression at nearly all levels of society; in schools, in business, in gangs and clubs, etc. It was only natural that emphasis on teamwork should appear at many places in the armed forces. Probably teamwork and team spirit were nowhere more emphasized than in the combat crews of the AAF. This emphasis began at the highest levels of command where publicity releases, recruiting drives, training policies and the like had their origin and extended down to the training organizations where the concept of the team was utilized to build morale and motivate individuals and crews in training for combat. One of the main reasons given for the sweeping changes in procedures for crew assembly which were directed in AAF Letter 50-117 was the belief that previous practices for assembling lead crews resulted in "breaking up the integrity of the combat crew" by taking individuals from separate crews and recombining them during their tour of combat duty. Whatever the ultimate evaluation of this emphasis, there is no doubt that the crew was quite generally treated as if it had a character quite apart from the individual characteristics of its members.

In attempting to evaluate the performance of the crew as a whole, one immediately meets an interesting and important problem in Social or Group Psychology. The fundamental question is: Are the characteristics of a

group of individuals, such as a combat crew, completely explained by summing up the individual characteristics of its members? From the official emphasis upon crew integrity, from interviews with returnees and the like it would seem that there is widespread belief that the crew as a whole does present characteristics not observable in the individual crew members. However, there were many with the opposite viewpoint who maintained that the one important thing in assembling a lead crew was to obtain individual crew members that were of outstanding ability in their specialties. Further evidence here was the fact that it was extremely difficult to obtain descriptions of crew characteristics, most of the suggested characteristics of good lead crews being stated in terms of characteristics of various individual crew members. Any evidence that might be obtained in studies of crew proficiency which could throw light on this problem would be of great importance for all branches of social science. It would also have important implications for general Air Forces policy.

It was of course difficult to obtain any direct evaluation of lead crews as such in operational training since few crews were given an opportunity to serve in true lead capacity. Furthermore, from the description in previous paragraphs it can be seen that most of the characteristics thought important in lead crews could be measured objectively only with great difficulty. Thus research personnel attempting to make evaluation of potential lead crew proficiency were forced to fall back upon subjective judgments. Finally, the importance attributed by all concerned to over-all proficiency of both individual and crew made any measures or observations of crew proficiency pertinent to evaluation of lead crew proficiency.

CRITERIA OF CREW PROFICIENCY

Measures of Bombing Accuracy

In a sense, the bombing accuracy of a crew has an inherent validity not possessed by any other criterion of crew proficiency, for the accurate bombing of enemy targets is the ultimate objective of all bombardment operations. Two main types of evidence were available in operational training relative to the bombing accuracy of all types of bombardment crews. These were the crew average circular errors, obtained from observations of practice bombs dropped on circular targets at bombing ranges, and the records of radar and camera bombing of industrial targets where no bombs were actually released. In heavy bombardment training greater emphasis was placed on the former type of operation and in very heavy bombardment training the emphasis was more on the latter type. Since rather different problems were encountered in studies of proficiency in the two types of operations, they are discussed separately in the paragraphs which follow.

Visual Bombing on Bombing Ranges

A number of studies were made in the Continental Air Forces of average circular error in operational training. The procedures ordinarily used in

obtaining and recording average circular error in operational training differed little from those used in Bombardier Schools in the AAF Training Command. In general, there were the same sources of error and the reliability of this measure was low, being approximately 0.50. However, this value is considerably higher than such values obtained in the AAF Training Command where they were frequently not significantly different from zero. This difference reflects a fundamental difference in procedure in that in the Training Command the pilot varied from mission to mission while in operational training the same crew and pilot flew with a bombardier throughout. Studies of bombing errors in the Training Command had shown that the pilot contributed about as much to bombing accuracy as did the bombardier. The combined effect of pilot and bombardier in operational training would thus be expected to produce a bombing error score of higher reliability than that in the AAF Training Command, as was actually found. The Training Command data do not permit of a sufficiently fine analysis of variance to determine whether or not the observed reliability in operational training was actually any different from what would be expected from a combination of the contribution of pilot and bombardier. Near the end of hostilities attempts were in progress to remove some of the sources of error in current practices by the installation of sonic scoring equipment and procedures. However, no data with the new procedures were ever obtained. For a more detailed analysis of circular error see chapter 7, Bombardier.

The most important criticism of average circular error as a criterion of crew proficiency was neither the lack of reliability nor the prevalence of sources of error in obtaining it. It was the criticism that the bombing procedures involved so little of the procedures used in combat operations. Average circular error was obtained from bombs dropped at *target ranges.* Hence the target location and appearance were well known to the crew. There were no problems of identification of target or IP. Frequently several runs were made in succession over the same target, and one or more practice runs were often made prior to dropping bombs for record purposes.

Simulated Bombing of Industrial Targets

For these reasons the main emphasis in training shifted to procedures paralleling combat operations more closely, namely those involving long range flights to industrial areas where a bombing run was made upon a target seen by the crew from the air for the first time. The chief problem in this type of training has been to obtain an accurate scoring of the bombing run.

Two main types of scoring of these bombing runs were developed. The first and most commonly used was camera scoring. There were a number of camera scoring systems but in general the techniques involved a succession of pictures taken on the approach, at the moment of hypothetical bomb release and perhaps a picture or two thereafter. By obtaining in

each photograph the point on the ground representing the true vertical from plane to ground it was possible to plot the flight path of the plane. If the interval of time between the photographs was accurately determined it was even possible to obtain from the photographs the ground speed of the plane and its altitude. The probable point of impact was then estimated and the circular error from the intended target determined. In practice a great many difficulties were encountered in the use of this method. But for some time it remained the only feasible method of scoring such bombing practice.

With the development of accurate radar ranging and directional equipment, it became possible to obtain a completely objective record of the flight paths of planes approaching industrial areas where such installations had been set up. Headquarters, Army Air Forces therefore initiated Project 584, Radar Bomb Scoring, to score all bombing runs on industrial targets by crews in training in the Continental Air Forces. Installations were not completed at the close of training so that very few data from this method of scoring are available.

Since detailed statistical data on both visual and radar bombing scores are presented in the discussion of average circular error in chapter 7, The Bombardier, none will be presented here. In general, radar bombing scores showed moderate reliability and from that standpoint were regarded as adequate for use as criteria in validation studies. In addition to moderate reliability, bombing scores, especially those obtained in VHB training from simulated bombing of industrial targets were believed to have special value as a measure of crew proficiency. It has already been pointed out in discussions of the use of bombing scores as criteria of proficiency for various different specialists, such as the pilot, bombardier, navigator and radar observer, that the performance of several crew members affected the bombing score as well as the proficiency of the specialist in question. As a result, the bombing scores were not highly regarded as measures of individual proficiency. For precisely this reason, however, bombing scores might be regarded as good measures of crew proficiency. While it may not be possible to determine the relative contributions of the different crew members, the bombing error score probably represents the best available measure of the combined efforts of the crew as a whole.

Rating Techniques

Data on potential lead crew proficiency were obtained in the Continental Air Forces with three different rating scales. Two of these, Scale A and Scale B, were developed in the 16th and 17th Bombardment Training Wings of the Second Air Force, while the third was developed by instructor personnel in the Lead Crew School at Muroc Army Air Field with the assistance of aviation psychologists in the Fourth Air Force. Scales A and B attempted to obtain an over-all judgment from the observer as to the amount or degree to which certain traits and characteristics were present in

the individual being rated. The scale developed in the Fourth Air Force was designed to rate student performance on a particular mission on various aspects of his job. With the Second Air Force scales, each crew member would be rated only once, near the end of training. With the Fourth Air Force scale each individual and crew would be rated on each of several missions. The rating procedures in the Second and Fourth Air Forces were thus very different even though some of the traits to be rated were similar.

The details of procedure used in constructing and administering Scales A and B in the Second Air Force are described in chapter 4, Airplane Commander, in the subsection on ratings. In addition to items referring to each aircrew officer and to the gunners of the crew as a group, a number of items in each scale referred to characteristics of the crew as a whole. Scale A contained six such items requiring ratings of the following: (1) Interphone Procedure (2) Crew Discipline (3) Calmness in Emergencies (4) Preflight Procedures (5) Eagerness and Enthusiasm and (6) Potential Lead Crew Proficiency. Scale B contained eight items including: (1) Potential Lead Crew Proficiency (2) Preflight Procedures (3) Eagerness and Enthusiasm (4) Interphone Discipline (5) Foresight and Planning (6) Teamwork (7) Crew Discipline and (8) Ranking in Proficiency (among other crews of the class). In addition, with Scale B confidential judgments were obtained from each airplane commander as to the five best and five poorest crews in his class. In table 11.2 are given the distributions of ratings on each item referring to the total crew for both Scale A and Scale B. Also given are the distributions of the total ratings on all the crew items for each scale. Note that with Scale B all crews were not rated on all items.

With both Scale A and Scale B independent ratings from two or more observers were obtained wherever possible at each of the stations where ratings were made. The reliabilities of such ratings are shown by the correlation between independent ratings of the same crews. These are given

TABLE 11.2.—*Distributions of ratings on items referring to the crew as a whole*

16th AND 17th WINGS—SECOND AIR FORCE

Scale A[1]—items numbered

Rating	Item numbers					
	1	2	3	4	5	6
1	35	43	27	54	52	[illegible]
2	149	134	161	202	145	128
3	117	133	128	63	107	133
4	18	9	4	1	15	9
5	1	1	0	0	1	1

[1] Two raters. 160 crews.

Scale B[2]—items numbered

Rating	Item numbers							
	1	2	3	4	5	6	7	8
1	54	62	49	40	42	53	66	67
2	76	105	87	46	100	92	114	91
3	93	95	114	60	92	71	57	102
4	57	31	44	11	32	23	25	32
5	14	5	4	3	2	1	10	8
Totals	294	298	298	160	268	240	272	300

[2] Number of crews varied.

Total score 6 items	*Number of ratings Scale A*
6	9
7	6
8	7
9	15
10	18
11	27
12	35
13	32
14	34
15	29
16	31
17	39
18	24
19	6
20	5
21	1
22	0
23	2
Total	320

Total score 8 items	*Number of ratings Scale B*
8–9	26
10–11	21
12–13	26
14–15	43
16–17	35
18–19	36
20–21	29
22–23	40
24–25	28
26–27	26
28–29	16
30–31	7
32–33	9
34–35	3
36–37	2
Total	347

in table 11.3 for each of the items of the scales. Coefficients were computed separately for each class or squadron rated. These were then averaged by Fisher's weighted z method to give the coefficients listed in the table. In table 11.4 are presented the reliabilities of the total scores for both scales, based on all items. Separate coefficients are given for each station and for all stations combined. It is clear that the reliabilities of the crew ratings were considerably higher for Scale B than for Scale A. It is also interesting that there is a fair correlation between the standard deviation of the ratings and the reliability, the lowest standard deviation for Scale B being higher than the highest for Scale A. From the distributions of the total ratings on all crew items in table 11.2 it is apparent that there is a

TABLE 11.3.—*Reliabilities of independent ratings of crew proficiency*

16th AND 17th WINGS—SECOND AIR FORCE

Item rated	Number of cases	First rater, 1		Second rater, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Scale A:						
1. Interphone	160	2.25	0.66	2.51	0.71	0.24
2. Discipline	160	2.26	.71	2.38	.69	.30
3. Calmness	160	2.33	.54	2.37	.51	.24
4. Preflight	160	2.04	.45	1.99	.48	.07
5. Eagerness	160	2.15	.60	2.32	.69	.28
6. Lead crew	160	2.36	.63	2.27	.72	.30
All items	160	2.24	.59	2.33	.57	.48
Scale B:						
1. Lead crew	147	2.72	1.18	2.61	1.07	0.71
2. Preflight	149	2.62	.73	2.41	.76	.48
3. Eagerness	149	2.75	.95	2.36	.96	.53
4. Interphone	80	2.36	.86	2.27	1.10	.26
5. Foresight	134	[illegible]	.99	2.54	.88	.44
6. Teamwork	120	2.21	.94	2.33	.92	.54
7. Discipline	136	2.41	1.08	2.11	1.01	.42
8. Ranking	150	2.17	1.04	2.43	.98	.64
All items	151	[illegible]	.82	2.35	.81	.68

good range of scores for the crew scales as a whole. Evidently the instructors did make some discrimination between the potential lead crew proficiencies of the different crews, although the basis of discrimination may have been spurious or irrelevant.

The reliability of the ratings for each different aircrew specialty rated are given in table 11.5 for Scales A and B. The coefficients are in each case based on the combined ratings on all items of the specialty. The

TABLE 11.4.—*Reliability coefficients of total scores of crew rating scales*

16th AND 17th WINGS—SECOND AIR FORCE

Station	N	First rater, *1*		Second rater, *2*		r_{12}
		Mean	S. D.	Mean	S. D.	
Scale A:						
Alamogordo 1	23	2.24	0.55	2.18	0.52	0.3[illegible]
Alamogordo 6	23	1.87	.38	2.05	.56	.[illegible]
Biggs	41	1.98	.58	1.96	.66	.3[illegible]
Clovis	36	2.73	.35	2.61	.38	.5[illegible]
Tucson	37	2.29	.38	2.72	.43	.11
Combined	160	2.24	.59	2.33	.57	1.[illegible]
Scale B:						
Great Bend	16	2.56	0.65	2.14	0.49	0.5[illegible]
Walker	50	2.55	.83	2.22	.89	.7[illegible]
Albuquerque	42	2.31	.96	2.22	.68	.5[illegible]
Pyote	43	2.56	.51	2.75	.79	.5[illegible]
Combined	151	2.48	.82	2.35	.81	1.0[illegible]

[1] Averaged by means of Fisher's z technique.

differences in reliability between Scale A and Scale B are not as marked in the case of the separate specialties as they are in the case of the ratings of the crew as a whole. In general, the scales show moderate reliability for all positions. It is interesting that the navigator and bombardier ratings, which have the lowest reliability in Scale B have the highest reliability in Scale A. It is probable that further study of the wording of the items on these scales would reveal the reasons for this outcome; however cessation of training prevented further investigation of this sort.

TABLE 11.5.—*Correlations between independent ratings*

16th AND 17th WINGS—SECOND AIR FORCE

Type of rating	Number of cases	First rater, *1*		Second rater, *2*		r_{12}
		Mean	S. D.	Mean	S. D.	
Scale A:						
Crew	160	2.24	0.59	2.33	0.57	0.4[illegible]
AC	160	2.41	.63	2.47	.55	.55
CP	160	2.71	.51	[illegible]	.54	.4[illegible]
N	160	2.76	.62	2.64	.60	.66
B	160	2.71	.68	2.62	.69	.74
FE	160	2.47	.90	2.52	[illegible]	.61
VO	160	2.50	.62	2.60	.66	.68
Gunners	160	2.68	.70	2.90	.72	[illegible]
Scale B:						
Crew	151	2.48	0.82	2.35	0.81	0.6[illegible]
AC	165	2.46	.84	2.54	.83	.75
CP	127	2.47	.71	2.40	.76	.54
N	92	2.63	.90	2.38	[illegible]	.46
B	142	2.54	.82	2.63	.79	.46
FE	116	2.86	.85	2.57	.82	.5[illegible]
VO	79	2.66	.79	2.46	.84	.59
ROM	143	2.33	.67	2.30	.90	.76
Gunners	93	2.62	.73	2.59	.83	.66

Item reliabilities for Scale A in all specialties ranged from 0.07 to 0.73 with a median coefficient of 0.36 (N of 160). For Scale B the coefficients ranged from 0.16 to 0.66 with a median of 0.41. (The number of cases per item varied in the case of Scale B from 52 to 151). There were differences between bases as to the reliability with which items were rated and considerable variation in reliability from item to item at a single base. The only general conclusion that seems warranted from the study of item reliabilties is that instructors showed greater agreement among themselves when rating students on general over-all items than when making ratings on specific items of behavior. The average coefficients of reliability for over-all or general items were 0.46 and 0.53 while those for the more specific items were 0.38 and 0.42 for Scales A and B respectively. In other words, instructors were more likely to agree that Student A was a good lead crew prospect than that he was good at his preflight check.

All scales for all aircrew positions showed high intercorrelations of items. This fact, taken together with the higher reliability of over-all items, indicates that there was considerable halo effect in the ratings on all scales.

The ratings of the crew were generally made by the same instructor who rated the airplane commander. As might be expected the ratings of the crew, in the case of both scales, showed a high correlation with the ratings of the airplane commander. Apparently, the instructor's rating of the crew was strongly influenced by his rating of the airplane commander. The appropriate correlations are given in table 11.6. All coefficients in the table represent the weighted z average of separate coefficients for each class and station.

The sum of the ratings of the crew as a whole showed only a low degree of correlation with the ratings given individuals other than the airplane commander. There was evidence of some correlation between ratings of copilots and navigators and crew ratings while the ratings of gunners were least related to the ratings of the total crew.

TABLE 11.6.—*Correlations between total score on crew scale and total score for each specialty scale*

16th AND 17th WINGS—SECOND AIR FORCE

Specialty	N	Crew scale, 1		Individual scale, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Scale A:						
1. Airplane commander	160	27.36	5.77	24.30	5.28	0.75
2. Copilot	160	27.36	5.77	21.61	3.99	.33
3. Navigator	160	27.36	5.77	21.67	4.37	.13
4. Bombardier	160	27.35	5.77	16.18	4.33	.06
5. Flight engineer	160	27.36	5.77	15.33	5.32	.24
6. Radar observer	160	27.36	5.77	20.35	4.77	.20
7. Gunners	160	27.36	5.77	11.27	2.45	.18
Scale B:						
1. Airplane commander	192	38.22	11.40	34.50	10.76	.92
2. Copilot	188	38.10	11.48	35.50	8.56	.57
3. Navigator	185	38.[illegible]	11.44	25.78	7.88	.23
4. Bombardier	186	38.[illegible]	11.44	36.14	9.88	.38
5. Flight engineer	186	38.30	11.36	35.98	9.92	.31
6. Radar observer	122	36.14	11.72	26.26	7.48	.33
7. Radio operator	150	37.88	11.60	23.58	6.56	−.02
8. Gunners	190	38.18	11.36	25.66	6.48	.28

The crew ratings obtained from the Lead Crew School at Muroc Army Air Field in the Fourth Air Force included ratings of three types. First, each crew member was rated on 6 to 10 specific items on each mission that was rated. Second, each crew member was rated on over-all proficiency on each mission. Finally, each crew was given an over-all rating as a potential lead crew on each mission. In table 11.7 are given the reliabilities of the individual item and over-all ratings. In obtaining these reliabilities an unweighted average individual rating was obtained from the item ratings. Then for all three types of ratings odd-even mission correlations were computed separately for each sample of crews. These coefficients were then corrected by the Spearman-Brown formula and the weighted average of the corrected coefficients for three classes at the school were obtained. The latter form the data included in the table. The over-all ratings of radar observer and crew and the average individual item ratings of the radar observer have a high degree of reliability. In general the other reliabilities are moderately high and roughly comparable to those obtained in the Second Air Force. One marked difference was in the navigator ratings which were considerably less reliable in the Fourth Air Force data. It is known that administrative difficulties were adversely affecting the reliability of the navigator rating situation at Muroc and the significance of this result is not clear from the data at hand. Distribution statistics are not presented since in all computations the ratings were converted to standard scores with a mean of 50 and a standard deviation of 10.

TABLE 11.7.—*Reliabilities of mission ratings of B–29 lead crews*

LEAD CREW SCHOOL, MUROC, FOURTH AIR FORCE

Position	Number of cases	r_{tt}
Average individual rating:		
Airplane commander	7[illegible]	0.45
Navigator	9[illegible]	.41
Bombardier	93	.[illegible]
Radar observer	85	.6[illegible]
Over-all rating:		
Airplane commander	70	.47
Navigator	98	.27
Bombardier	93	.49
Radar observer	85	.75
Crew rating	102	.74

Crew Proficiency Check

The Crew Proficiency Check (VHB) developed in the Third Air Force represented a somewhat different approach to proficiency measurement. The purpose of the instrument was to obtain as objective a measure as possible of over-all crew proficiency. Surveys of measures of crew proficiency had revealed none that were reliable. It was, therefore, deemed necessary to construct a special instrument for the purpose. Findings of previous research formed the basis of much of the developmental work. Most of the background material was taken from studies on the qualifications of an airplane commander and evaluation of airplane commnder

ability in CCTS training done in the Third Air Force. These studies are presented in detail in chapter 4, Airplane Commander. In these studies descriptions of combat situations in which airplane commander ability was exhibited were analyzed to obtain information on the qualifications of successful airplane commanders. Research personnel then flew with student crews on nearly all of the required missions in CCTS training keeping logs on each mission, in which record was made of all observed aspects of briefing, preflight, flight, and post flight activities of airplane commander and crew. These logs were analyzed to determine those areas or activities where airplane commander ability was manifested and where measurement could be introduced.

During these studies it was also possible to obtain information on the extent to which the airplane commander contributed to the proficiency of the crew as a whole. Since the airplane commander was the individual around whom all crew activities revolved, since he was largely responsible for much of the training of the crew, and since he was immediate superior officer of all crew members, it was not surprising to observe that the proficiency of the individual airplane commander was reflected to a large extent in the proficiency of the crew as a whole.

From analyses of qualifications of airplane commanders obtained from combat situations, from analyses of flight logs, and from general observations and discussions with supervisory and instructor personnel, an airplane commander check list was constructed. Developmental work was carried on until the program for the selection of potential lead crews was inaugurated. Because of the relationship between airplane commander ability and over-all crew proficiency, the airplane commander check was modified to become a crew proficiency check for very heavy bombardment crews. Airplane commander items which did not contribute to over-all crew proficiency were dropped and items reflecting proficiency of other crew members, intracrew relations and cooperation were introduced.

The Crew Proficiency Check (VHB; B–29) in its final form consisted of a crew history cover-page and 53 items arranged in sequence to follow the logical progression of a mission from briefing to critique. This item arrangement was made to facilitate the administration of the check list since it had to be done during the mission. Areas covered by the items were: General Proficiency; Discipline, Air and Ground; Intracrew Cooperation; and Crew Compatibility. The items were largely of the multiple-choice, descriptive-objective type. Each item dealt with an area of behavior or with specific acts of crew members with the multiple-choice alternatives being the possible reactions or ways of meeting the problem involved. Some items involved general over-all ratings of crew or crew members.

A preliminary form of the check was administered to 40 crews at Barksdale Field in July 1945. From an analysis of this administration further refinements were introduced and the final form was administered to 130 crews in training at four Third Air Force CCTS stations. The check was

administered to all crews with whom an instructor pilot rode during the period 5-15 August 1945 regardless of mission or phase of training. Two scoring formulas were developed. One assigned weights to alternatives according to judgments of "goodness" or "poorness" of the alternative responses, determined in conferences with training personnel and from study of training directives and the like. A more refined scoring procedure yielded scores that correlated 0.94 or more with the method described and it was, therefore, discarded.

In table 11.8 are given the means and standard deviations of scores obtained at each station and the same data for the total group of 130 crews. Cessation of hostilities and the subsequent curtailment of training prevented the collection of sufficient data for determining the reliability of the Crew Proficiency Check (VHB).

TABLE 11.8.—*Scores on crew proficiency check (VHB; B-29)*

B-29—THIRD AIR FORCE

Station	Weeks in training	Number of crews	Mean score	S.D.
MacDill	1–10	17	30.94	7.20
Gulfport	1–4	25	31.48	4.85
Do	5–7	25	33.96	8.27
Chatham	1–4	19	33.26	7.04
Do	8–10	18	34.83	8.74
Barksdale	1–10	26	32.96	7.34
Total group	....	130	32.93	7.40

Rate of Mission and Training Accomplishment

Another type of criterion which initially appeared to be promising was the efficiency and dispatch with which crews attacked and completed their training requirements. Various training accomplishment scores were studied by research personnel. For detailed discussion of this type of criterion and attempts to measure its reliability see chapter 4, Airplane Commander.

In one of the studies Second Air Force research personnel obtained certain accomplishment-of-training scores for all crews on each of four classes (VH) for whom rating scale scores had been obtained for all crew personnel. Two training accomplishment scores were arrived at for each crew. One score was a measure of the number of Second Air Force training requirements that had been completed at the end of a given period of training while the other was based on the number of Army Air Forces Training Standard requirements completed. In the case of both scores, a mission or requirement flown but not completed was counted as one point. When a mission or requirement was completed it was counted as two points. It was found that the degree of correlation between the two measures of training accomplishment was quite high, as might be expected. At the end of 6 weeks of training the coefficient was 0.83 for 170 crews. At 12 and 18 weeks the 2 measures correlated 0.82 and 0.74 respectively, based on data from 90 crews.

The advantages and disadvantages of this type of criterion are discussed

at length in chapter 4, Airplane Commander. It will suffice here to repeat that no satisfactory measure of reliability is available for this criterion and it is known to have been influenced by a large number of variable factors for which controls are difficult if not impossible. Crew accomplishment scores apparently had some degree of reliability, but not much more can be said with certainty.

Crew Awards

Several stations in the various air forces made an attempt to reward crews that did an outstanding job during training. The rewards took various forms. In some cases rather valuable prizes were given. At several stations winning crews were given the privilege of flying to the home city of some crew member on a short leave or furlough. In other cases the honor of the citation was considered sufficient reward. Research personnel felt that the selection of outstanding crews for these awards might provide a criterion of proficiency that would be useful in validation and lead crew studies.

Examination of the award procedures indicated that crew awards had little value as a criterion of crew proficiency. For example, at one station in the Second Air Force, one crew was chosen each week as the "Crew of the Week." Over-all ratings made by training personnel of the individual officers of four successive crews of the week were examined. Two of these crews had officers with very high ratings. In the case of the other two, several of the officers had almost the lowest possible ratings. Inquiry revealed that the Director of Training thought that these crews needed a "shot in the arm" to improve their morale. Hence he had had them designated as crew of the week.

At one station a fairly standard procedure had been set up to determine the ten most successful crews in training to whom awards would be made at the end of training. Since records used in this determination had been kept on file for several recent classes, a more detailed study of this criterion was undertaken. The scores entering into the crew award score together with the intercorrelations between them are shown in table 11.9. The award score was obtained simply by adding together all of the other scores. From the intercorrelations it can be seen that the Military Training Grade determined to a large measure what the award score of the crew would be. This is due to the fact that the Military Training Grade had a considerably larger spread of scores than any other measure. Also given in the table are the correlations between the various grades and scores at the end of the first half of training and those for the total training period. These coefficients are spuriously high since the first half scores were combined with second half scores to obtain the total score.

Mission Failures

Comments on the unsatisfactory performance of certain crews sometimes included references to a tendency to return from flights without accomplish-

ing the intended missions. Such returns are usually called "abortions" and certain records were kept of their occurrence. Sometimes the abortions took the form of personnel failure and were so reported. But more often the reasons given for mission failure involved malfunctions of plane and equipment. Attempts to study this criterion met with the difficulty that the available records did not contain enough information to separate abortions that represented good judgment from those that were caused by poor motivation and mistakes of the crew.

TABLE 11.9.—*Reliabilities and intercorrelations of criteria used in crew awards*

COWEN FIELD, SECOND AIR FORCE

Type of score	Code[1]	A	B	C	D	E	F
Intercorrelations:							
Award score	A		0.12	0.56	0.8[illegible]	0.62	[illegible]
Crew academic average	B			−.63	.06	.20	[illegible]
Crew average flying grade	C				.12	.36	[illegible]
Crew average military grade	D					.32	[illegible]
Crew average aptitude rating	E						[illegible]
Average circular error	F						

[1] N = 112.

Type of criterion	Number of Cases	First half, 1		Total, 2		[illegible]
		Mean	S. D.	Mean	S. D.	
Estimates of reliability:						
Award score	56	87.5	3.6	91.5	1.5	0.73
Crew average academic grade	56	85.7	2.6	86.6	1.7	.36
Crew average flying grade	56	85.1	5.3	91.0	3.0	.66

Intercorrelations of Criteria

Table 11.10, 11.11, 11.12, and 11.13 present data that were obtained showing the relationships between various measures of crew proficiency. First of all, ratings of the crew as a whole on various rating scales showed no relationship either to practice or radar bombing scores. Since the ratings and the bombing scores both showed moderate reliability this result suggests that there was real independence in the two measures. It should be pointed out in this connection that the instructors making the ratings of the crew as a whole did not consult the bombing scores and, at least in the case of radar bombing scores, did not have access to them. Similar evidence is found in the fact that the correlation between crew accomplishment of training at the end of 6 weeks and the practice bombing average circular error of the bombardier was 0.01 for 974 crews (B–17 and B–24) in the Second Air Force.

On both Second Air Force Descriptive Rating Scales crew ratings showed only a slight relationship to crew progress as indicated by the number of training requirements completed at a given stage of training. This was true for both of the training accomplishment scores: number of missions completed and AAF Standards completed.

TABLE 11.10.—*Correlation between practice ACE and crew ratings, Scale A*

B-29 CREWS—SECOND AIR FORCE

Station	N	Circular error, *1*		Crew rating, *2*		r_{12}
		Mean	S. D.	Mean	S. D.	
Alamogordo 6	23	339.90	63.40	26.48	5.16	0.12
Alamogordo 3	23	330.50	72.08	23.28	5.32	−.08
[illegible]	41	256.19	59.60	23.62	5.14	−.12
Clovis	36	241.76	46.40	32.03	3.83	.04
Tucson	37	231.95	59.40	29.94	3.43	.20
Combined (average by Fisher's z)	160	254.13	69.16	27.36	5.85	.08

TABLE 11.11.—*Correlations between crew rating scale scores and radar circular error T-scores*

B-29 CREWS—SECOND AIR FORCE

Station	N	Crew rating, *1*		Circular error, *2*		r_{12}
		Mean	S. D.	Mean	S. D.	
Scale A:						
Alamogordo 3	11	27.73	4.02	49.73	5.14	0.04
Alamogordo 6	16	21.62	3.63	50.00	6.53	−.25
Clovis	35	32.09	3.87	49.36	3.52	.03
El Paso	23	23.64	5.25	50.14	5.04	.13
Tucson	35	29.90	3.50	49.24	5.42	.07
Combined (average by Fisher's z)	125	27.56	5.83	49.62	5.02	−.02
Scale B:						
Great Bend	31	35.90	12.08	50.11	5.32	0.03
Walker	34	41.02	12.32	49.25	4.24	−.33
Albuquerque	21	30.18	9.35	50.02	4.60	.02
Combined (average by Fisher's z)	86	36.54	12.32	49.99	4.66	−.12

TABLE 11.12.—*Intercorrelations of measures of crew proficiency*

B-29 LEAD CREW SCHOOL—FOURTH AIR FORCE

Variables[1]	1	2	3	4
1. ACE total	.....	0.74	0.87	0.15
2. ACE normal aiming[2]	.....	.....	.45	.18
3. Percent gross errors[2]	.....	.....	.....	.09
4. Over-all crew rating	.....	.....	.....	.....

[1] N = 102.

[2] For a discussion of normal aiming error and percent gross errors see chapter 7, Bombardier.

TABLE 11.13.—*Correlations between descriptive ratings of combat crews and training accomplishment scores at six weeks*

SCALES A AND B—SECOND AIR FORCE

Scale	N	Total rating, *1*		Missions accomplished, *2*		r_{12}
		Mean	S. D.	Mean	S. D.	
No. of missions accomplished:						
Scale A		2.27	0.48	36.8	5.6	0.18
Scale B		2.37	.72	32.8	12.3	.25
Number of training standards completed:						
Scale A		2.27	.48	38.7	17.0	.19
Scale B		2.37	.71	36.2	18.6	.19

VALIDATION OF INDIVIDUAL STANINES AGAINST CREW CRITERIA

Average Circular Error

For data on the relation between the stanines of individual crew members and the average circular error of the crew, see Chapter 4 through 10. In general, very low or zero correlation was found. The interpretation of this lack of relationship has varied, depending on the specialty concerned. In some cases, there might well be no relationship between the stanine of the specialist and bombing scores. (In other cases, a possible relationship might have been obscured by the fact that neither of the scores is completely reliable).

Crew Ratings

The augmented pilot stanines of 73 airplane commanders in the 16th Wing of the Second Air Force correlated 0.11 with the rating of the crew. In the case of Scale B (17th Wing), coefficients were 0.08, —0.08 and 0.04 for the bombardier, navigator and augmented pilot stanines of 86 airplane commanders. Ratings of the crew were not compared with the stanines of the other aircrew specialists.

Crew Awards

The correlation between stanines of various crew members and certain scores used in making crew awards are given in table 11.14. None of the correlations is significantly greater than zero.

TABLE 11.14.—*Correlations between stanines and crew award scores*

SECOND AIR FORCE

Type of score	Number of cases	Stanine, 1		Criterion, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Airplane commander's pilot stanine against:						
Crew award score[1]	189	6.54	1.6	50.5	9.4	−0.11
Crew academic average	189	6.54	1.6	84.5	1.9	−.0[illegible]
Crew flying average	189	6.54	1.6	89.5	4.6	.0[illegible]
Crew military average	189	6.54	1.6	92.3	8.4	.01
Crew attitude rating	189	6.54	1.6	2.0	.3	.02
Copilots pilot stanine against:						
Crew award score[1]	211	6.04	1.7	50.0	9.8	.07
Crew attitude rating	211	6.04	1.7	2.0	.4	.14
Navigator's navigator stanine against:						
Crew award score[1]	192	7.65	1.1	50.4	10.2	−.04
Crew attitude rating	192	7.65	1.1	2.0	.3	.03
Bombardier's bombardier stanine against:						
Crew award score[1]	212	6.12	1.[illegible]	50.4	10.1	−.04
Crew attitude rating	212	6.12	1.[illegible]	2.0	.04	−.05

[1] After total crew award score had been prepared, crews were ranked according to this score and the ranks were converted to standard scores with a mean of 50 and S.D. of 10.

Rate of Training Accomplishments

Data on the relation between training requirements completed at 6 weeks and stanines of various crew members are given in chapters 4 through 10. Airplane commander stanines were found to correlate significantly with

training accomplishment scores. None of the stanines of other crew members were related significantly to training accomplishment.

VALIDATION OF COMBINATIONS OF STANINES AGAINST MEASURES OF CREW PROFICIENCY

Introduction

Many of the criteria discussed earlier in this chapter reflect the ability of several, and in some cases, of most of the members of the crew. It seemed worthwhile, therefore, to determine whether a score, representing more or less the average aptitude score of the crew members was related to these measures of proficiency. A number of such comparisons were made by research personnel in the Second Air Force. Since stanines were not available for all of the commissioned personnel, it was necessary to make separate studies with a number of different groups; i.e., a group of crews with stanines for 4 specialties, crews with stanines for 5 specialties, and crews with stanines for all aircrew officers. In all studies, the augmented pilot stanine was used for both airplane commander and copilot. For the flight engineer the nonaugmented pilot stanine was used. For the other aircrew positions the appropriate stanine was used: N stanine for navigator, B stanine for bombardier and N stanine for radar observer.

Crew Ratings

The correlations between the average stanines of crew members and ratings of the crew as a whole on both scales are shown in table 11.15. In

TABLE 11.15.—*Correlations between average crew stanines and total score on crew rating scales*

16th AND 17th WINGS—SECOND AIR FORCE

	N	Stanine, 1		Crew rating, 2		r
		Mean	S. D.	Mean	S. D.	
Crew rating scale A:						
Average six stanines	58	6.97	55	29.23	4.86	0.07
Average five stanines	66	6.83	54	27.61	5.43	−.03
Average four stanines	28	6.88	59	24.22	5.10	.23
Average (regardless of number)	159	6.86	54	27.11	5.64	[1].04
Crew rating scale B:						
Average stanine (Great Bend)	46	7.27	0.92	36.02	11.08	.20
Average stanine (Walker)	50	6.86	1.02	38.46	12.40	−.18
Average stanine (Albuquerque)	45	6.60	0.63	35.06	10.72	−.09
Average stanine (Pyote)	50	6.89	0.62	42.62	9.68	.18
Combined	191	6.91	0.25	38.14	11.40	[1].01

[1] Average by Fishers z of separate coefficients for different stations and classes.

table 11.16 are shown the validities for comp ite stanines when the total scores on all the individual rating scales were added together to form a crew rating. Extensive work was done with thes data, combining both rating scores and stanines in several different wa All combinations dem-

onstrated the same lack of relationship as is shown in tables 11.15 and 11.16.

TABLE 11.16.—*Correlations between average crew stanines and total of scores of all individual rating scales*

16th AND 17th WINGS—SECOND AIR FORCE

	N	M_1	SD_1	M_2	SD_2	r_{12}
16th Wing scale A:						
Average six stanines	58	6.97	.55	134.88	13.47	0.0[illegible]
Average five stanines	66	6.83	.54	128.80	14.08	.0[illegible]
Average four stanines	28	6.88	.59	126.20	12.60	*.41
Average (regardless of number)	159	6.86	.56	129.24	14.61	[1].13
17th Wing scale A:						
Average stanine (Great Bend)	46	7.27	.92	4.91	.90	−0.0[illegible]
Average stanine (Walker)	50	6.86	1.02	4.83	.85	−.20
Average stanine (Albuquerque)	49	6.58	.67	5.35	.79	−.01
Average stanine (Pyote)	50	6.89	.62	5.35	.79	*.29
Combined	195	6.90	.86	5.11	.87	[1].01

[1] Average by Fisher's z of separate coefficients for each class and station.

Crew Awards

There was in general little relationship between crew average stanine and scores used in making crew awards at Gowen Field in the Second Air Force, as is evident from table 11.17. However, two of the coefficients are significant at the 5-percent level. In this connection, it should be pointed out that the attitude rating and average academic grade were highly correlated and both scores represent much the same accomplishment. As with validity studies of individual stanines, there was some prediction of academic success by the stanines.

TABLE 11.17.—*Correlations between crew average stanines and crew award scores*

GOWEN FIELD—SECOND AIR FORCE

Type of crew award score	Number of cases	Average stanine, 1		Crew award, 2		r_{12}
		Mean	S. D.	Mean	S. D.	
Total award score	158	6.56	0.7	88.2	4.8	0.01
Average attitude rating	158	6.56	.7	2.2	.2	*.15
Average academic grade	158	6.56	.7	84.6	1.8	*.14
Average flying grade	158	6.56	.7	89.4	4.7	−.03
Average military grade	158	6.56	.7	91.3	10.1	.01

Training Accomplishment

The correlations between various types of training accomplishment scores and crew average stanines are given in table 11.18. Although two of the correlation coefficients are significant with the large number of crews involved, there is little evidence of a practical degree of relationship. The stanines of the airplane commander showed a much higher degree of relationship to this criterion than does the crew average stanine. It would seem that the rest of the crew played a small part in connection with this criterion.

TABLE 11.18.—*Correlations between crew average stanines and training accomplishment*

SECOND AIR FORCE

Period of training	Crew stanine v number of missions completed	Crew stanine v missions completed per hour flown
215 B-17 crews, Ardmore:		
Fourth week, all phases	0.01	0.07
Sixth week, all phases	.14	*.16
Eighth week, all phases	.05	.06
Fourth to sixth week, all phases	.10	....
Sixth to eighth week, all phases	−.23	....

	Crew stanine v. requirements completed
974 B-17 and B-24 crews, many stations:	
Phase I, sixth week	0.08
Phase II, sixth week	.10
Phase III, sixth week	.09
Requirements, sixth week	.07
Total score, sixth week	.10

	Number of crews	Crew stanine v 2 AF flying training requirements completed	Crew stanine v AAF standards completed
B-29 crews, 16th and 17th Wings:			
Sixth week	223	0.03	0.02
Twelfth week	90	.07	.08
Eighteenth Week	90	.16	.07

RELATIONS BETWEEN PROFICIENCY OF DIFFERENT CREW MEMBERS

Assembly of crews in the Continental Air Forces was done for the most part with regard only to military occupational specialty and certain administrative requirements. Some attention was paid to relative ages, ranks and the like, but little or no effort was made to equate crew members in terms of proficiency. It would, therefore, be logical to assume that at the outset of operational training there is only a chance relation between the proficiency of the different crew members. However, the hypothesis has occasionally been advanced that in the process of living and working together crews tend to become increasingly homogeneous in many ways, even with respect to proficiency. A number of studies were made to test this hypothesis.

In the Fourth Air Force three classes of heavy bombardment crews from three stations were studied to see what relationships there were between the available proficiency measures for different crew members. The available measures of proficiency were converted to standard scores and the intercorrelations among the measures for the four aircrew officers determined. These are given in table 11.19 for the 1[illegible]0 crews studied. Also shown are the mean and standard deviations of the intercorrelations and the mean intercorrelations for each station for each [illegible] three types of profi-

ciency measures. The mean intercorrelations of synthetic trainer scores and ground school grades are approximately zero while those obtained for instructors' ratings are positive and significantly different from zero. This result might indicate that there was some homogeneity within crews although other types of measures do not support this view. On the other hand, the result might indicate the presence of a halo effect, i.e., the individuals of the crews were rated in conformity with a general impression made by one or more of the crew members.

TABLE 11.19.—*Intercorrelations among measures of proficiency of different specialties*

B-24 TRAINING—FOURTH AIR FORCE

All bases combined ($N = 140$)

Criteria	1	2	3	4	5	6	7	8	9	10	11
Airplane commander:											
Over-all ability, 1		.14	0.32	−0.01	0.19	0.08	0.18	0.21	−0.07	0.19	−0.06
Link trainer, 2			.07	−.04	.22	.87	.12	.11	−.02	.38	.13
Copilot:											
Link Trainer, 3				.14	.08	−.03	.11	.21	.05	.02	−.06
Engineer test, 4					.19	.02	.13	.18	−.09	.00	−.11
Navigator:											
Over-all ability, 5						.26	.19	.38	−.05	.07	.16
CNT rating, 6							−.11	−.03	.12	−.11	−.16
Ground school rating, 7								−.14	.06	.04	−.37
Bombardier:											
Over-all ability, 8									.15	−.01	.06
ACE[1], 9										.01	.02
7-A-3 rating[1], 10											.08
Ground school[1] average, 11											

[1] Signs of coefficients have been reversed to indicate positive association of goodness of performance.

Mean intercorrelations among proficiency measures shown above

	Muroc	March Field	Walla Walla
Mean of coefficients	0.04	0.05	0.10
S. D.	.13	.14	.17
Number of coefficients	.24	.44	.44

Mean intercorrelations according to type of measure

Criteria	Muroc	March Field	Walla Walla	All stations
Over-all ratings	0.33	0.25	0.23	0.26
Trainer ratings	−.01	−.02	.08	.02
Ground school grades		−.05	.15	.05

The correlations between rankings of different crew members by Second Air Force instructors are given in table 11.20. A total of 11 CCTS (H) classes are represented in the data. Except for the high correlations between the rank given pilots and copilots the correlations between ratings of different crew members are low. The similarity in ranking of pilot and

TABLE 11.20.—*Intercorrelations among rankings of different crew members*

11 B-17 AND B-24 CCTS CLASSES—SECOND AIR FORCE

Crew member		A	B	C	D
Airplane commander	A		0.65	0.10	0.18
Copilot	B	439		.16	.13
Navigator	C	602	437		.03
Bombardier	D	602	437	602	

GENERAL NOTE—Entries below the diagonal are the number of cases for the corresponding correlation coefficients.

copilot is probably due to the fact that the latter were not well known to the instructors. Such correlations as were here obtained can probably best be explained as due to a generalized halo effect in the minds of the raters. In fact, in view of the emphasis on total crew in operational training, these correlations are surprisingly low.

The intercorrelations between the scores used in making crew awards for different crew members are listed in table 11.21. As in the studies in the Fourth Air Force there is little evidence of homogeneity of proficiency within crews except when ratings are used as the measure of that proficiency.

TABLE 11.21.—*Intercorrelations of crew award scores of different crew members*

GOWEN FIELD—BOISE, IDAHO

Criteria[1]		A	B	C	D
Academic average:					
Airplane commander	A	......	0.02	−0.04	−0.13
Copilot	B	......	......	.06	.09
Navigator	C	......	......	......	.10
Bombardier	D	......	......	......	......
Attitude rating:					
Airplane commander	A	......	−.21	.31	.13
Copilot	B	......	......	.09	.03
Navigator	C	......	......	......	.09
Bombardier	D	......	......	......	......

[1] N = 112.

Intercorrelations of the over-all and average ratings of individuals and the rating of the crew, in the case of the Muroc scales, provide some insight into the relative contributions of the various specialists to crew proficiency (at least insofar as crew proficiency was measured by the rating of the crew). In contrast to the crew ratings in the Second Air Force, at the Lead Crew School every instructor rated the crew as a whole, regardless of his specialty. In the crew over-all rating, therefore, were represented, in about equal proportions the ratings of the crew by instructors in all specialties except flight engineer and gunnery instructors. Examination of the coefficients in table 11.22 indicate that the ratings of the pilot, navigator and radar observer were a little more closely related to the ratings of the crew than were the ratings of the bombardiers. Since this was a school for

TABLE 11.22.—*Correlations of individual rating scores with rating of crew and with radar ACE*

LEAD CREW SCHOOL, MUROC, FOURTH AIR FORCE
(N = 102)

Criteria	With over-all crew rating	With total radar ACE
Average rating of pilot	0.47	0.04
Over-all rating of pilot	.52	−.01
Average rating of navigator	.60	.19
Over-all rating of navigator	.45	.30
Average rating of bombardier	.35	−.01
Over-all rating of bombardier	.46	.09
Average rating of radar observer	.51	.19
Over-all rating of radar observer	.56	.20
Average rating of flight engineer	.01	.08
Over-all rating of flight engineer	−.08	.08

bombing, it is interesting that the bombardier contributed so little. On the other hand, it is not surprising that the ratings of flight engineer were not related to over-all crew rating since flight engineer instructors did not participate in the ratings of the crew.

Examination of the correlation of the individual ratings with radar average circular error scores, also shown in table 11.22 suggests that the performance of these two specialists as measured by the rating scale was somewhat more closely related to crew performance as measured by the bombing score than was the case for the other aircrew positions.

THE RELATIONSHIP BETWEEN INTERESTS AND MEASURES OF LEAD CREW PROFICIENCY

Although it was recognized as important, this problem was not exhaustively studied by research personnel due to the cessation of training at the end of the war. However, some preliminary studies were carried on. Since being a member of a lead crew involved certain hazards and responsibilities in addition to certain rewards in the way of prestige, it was reported by training officials that not all individuals were interested in being a member of a lead crew. Hence it was believed that the interest of individuals in the crew in being part of a lead crew would have some bearing on the efficiency with which they carried on their duties and the manner in which the crew as a whole performed.

Research was done in this area in both the 16th and 17th Wings of the Second Air Force with the descriptive rating scales that were described earlier. In the 16th Wing each individual was asked to indicate his strength of interest in being a member of a lead crew and his appraisal of the amount of his crew's interest in being a lead crew, and several other questions varying slightly in approach and wording to determine the strength of interest in lead crew membership. In the 17th Wing, the airplane commander was asked to estimate his crew's desire for a lead assignment and each crew member was requested to indicate his interest in being a member of a lead crew.

In the 16th Wing the correlation between the sum of the interests of all crew members in being members of a lead crew and the ratings of the crew as a potential lead crew was low (0.31) but significantly different from zero. This was obtained for 156 crews in the 16th Wing. For this same group, the correlation between the individual's interest in lead crew and his rating as potential lead crew material was not different from zero for airplane commanders, copilots, bombardiers and navigators. Low but significant correlations were found for radar observers and flight engineers. Exactly the same result was obtained when the interests of the individuals in having their crew be a lead crew were correlated with their rating as potential lead crew specialists.

As might be expected there was a high correlation in the case of all aircrew positions between the strength of the individual's desire for lead crew

duty and his estimate of the strength of the desire of the crew for that duty. This might be due to an actual relationship or to the fact that the individual projected his desires on the crew. Because of this high correlation, estimates of individuals of the strength of the interest of their crew for lead crew duties showed the same relationship to ratings of the crew and of individuals as was described above for strength of personal interest.

A low but significant relationship was found between the desires of the airplane commander and those of other crew members for lead crew duty. There was no evidence that personnel of one aircrew specialty were more desirous than those of any other for lead crew assignment. All personnel, regardless of position, reported on the average they were highly interested in lead crew assignments.

In the 17th Wing each crew member, except the airplane commander, indicated on a 9-point scale the strength of his interest for lead crew assignment. The airplane commander was asked to estimate on a similar form his crew's desire for lead duty. The correlation coefficient between the above estimates and instructors ratings of proficiency of airplane commanders on Scale B was 0.29. Correlation coefficients between crew members preferences and instructors ratings of respective crew members were: 0.25 for 148 copilots; 0.17 for 155 navigators; 0.30 for 147 bombardiers; 0.18 for 143 flight engineers; 0.13 for 71 radar observers. While low, these coefficients indicate that some relationship existed between the two measures. The correlation coefficient between the average interest of all crew members except the airplane commander and the airplane commanders' estimates of their crews' desires for lead duty was 0.29.

On the basis of the above results, conclusions regarding the relationship between interest in lead crew assignment and measures of proficiency must be highly tentative. It does appear, however, that the two factors are related in spite of the inadequacy of the criteria. The preliminary research suggests further that while preferences of the airplane commander were related to the interests of the other crew members, they were not major factors in determining those interests. The level of interest expressed in lead crew assignment was high throughout with no specialty showing greater interest than any other. The fact that the ratings of interest were bunched near the high end of the scale probably lowered the relationship between interest and proficiency.

STUDIES IN THE ASSEMBLY OF CREWS

Problem

Two important characteristics of good lead crews have implications with respect to assignment of individuals to combat crews. These are proficiency (or ability) and teamwork. With the increasing importance of lead crews the question was: given a particular group of individuals to be assigned to combat crews, how should they be assigned so as to produce the maximum number of outstanding crews? A number of factors were suggested for con-

sideration in the matching of individuals for assignment to particular crews. AAF Letter 50-117 specified some of these that had to do primarily with proficiency. Teamwork, however, depends greatly upon how well crew members get along with each other, how compatible they are. Among the factors suggested as important for compatibility were age, rank, geographical origin, hobbies and interests, and special friendships. It was the opinion of training personnel that individuals that were similar in terms of these factors would be more compatible than individuals that had extreme differences in one or more factors. Some study was, therefore, made of the role of these factors in the assembly of crews.

Evaluation of Crew Assembly in the Second Air Force

Routine procedure for the assignment of individuals to crews in the Third Air Force up until the fall of 1944 paid little attention to any factors that might improve the quality of the crews in terms of proficiency and compatibility. Assignment was usually based solely upon military occupational specialty and administrative convenience. In the Second Air Force, however, processing personnel at the combat crew classification and routing pool reported that an attempt was made to secure crews with good morale by sorting the trainees into crews on the basis of compatibility. For example, the formal structure of the crew requires that the airplane commander be in charge. An attempt was made to assign to his crew other crew members who would accept him in that status without question. Thus when possible, the airplane commander was the oldest and highest ranking officer on the crew. When an officer or EM waiting assignment requested assignment with some one or two other officers or EM, this was done if it did not interfere with the meeting of the quotas. Also considered to a lesser extent were geographical background and experience. No conscious attempt was made to consider proficiency or ability, although, it was well recognized that the positions of airplane commander and copilot required different levels of ability. However, the decision regarding assignment to one of the two positions was usually made in the AAF Training Command prior to arrival in the Second Air Force.

From data available in other connections, the ages and stanines of aircrew officers were obtained for approximately 24 CCTS (H) classes in the Second Air Force. These data were analyzed to determine the extent to which the above factors were considered in assigning individuals to crews. In table 11.23 are given the means and standard deviations for the stanines and ages of the occupants of four aircrew positions for which such data were available. The data show that the airplane commanders tended to be slightly older than the other officers. The differences in mean stanines (except for airplane commander and copilot) reflected largely differences in standards of admission to various types of cadet training. For example, requirements for navigation training had long been higher than for the

other specialties. Since the assignment of pilots as airplane commanders and copilots was carried out in the AAF Training Command supposedly without knowledge of the stanines, the difference in stanine between the two would seem to provide additional evidence of the validity of the stanine.

TABLE 11.23.—*Ages and stanines of aircrew officers*

24 B-17 AND B-24 CCTS CLASSES—SECOND AIR FORCE

Duty specialty	Stanines		Ages		Number of cases
	Mean	S. D.	Mean	S. D.	
Airplane commander	6.24	1.65	22.8	2.06	776
Copilot	5.39	1.62	[illegible]	1.92	[illegible]
Navigator	7.43	1.56	22.2	2.37	924
Bombardier	6.42	1.32	23.7	2.37	585

In table 11.24 are given the correlations between the stanines and ages of the airplane commanders and those of the other crew members. It can be seen that age was given considerable weight in determining assignments to crews, particularly as regards airplane commanders and copilots. The positive correlations between stanines, though low, are somewhat unexpected since aptitude scores presumably were not considered in crew assignments. The most obvious explanation would seem to be that this relation between stanines was brought about indirectly through the age factor. If the older and more experienced copilots were assigned to the same crews as the older and more experienced airplane commanders, two types of influences might operate to produce a correlation between their stanines. One would be whatever selective factors (either positive or negative) operated to delay older pilots and airplane commanders so as to arrive at the same time at the stage of operational training. The second type of influence would be any trends in population stanines over a period of time, with different pairs of airplane commanders and copilots being drawn from different testing populations.

TABLE 11.24.—*Correlations between ages and stanines of airplane commanders and those of other crew members*

24 B-17 AND B-24 CCTS CLASSES—SECOND AIR FORCE

Crew members	Number of cases	Correlation between stanines	Correlation between ages
Airplane commander v. copilot	785	0.21	0.51
Airplane commander v. bombardier	521	.16	.10
Airplane commander v. navigator	887	.02	.15
Airplane commander v. crew average	465	.14	.33

Crew Changes During Operational Training

Analysis was made of the crew changes occuring in two B–24 CCTS classes at one station in the Fourth Air Force and in five B–17 and B–24 classes at five stations in the Second Air Force. The frequencies of changes

in crew personnel are given in table 11.25. Although the percentages of crews with one or more changes was about the same in the two air forces, the percentage of individuals changed was considerably higher in the Fourth Air Force.

TABLE 11.25.—*Incidence of crew changes in operational training*

SECOND AND FOURTH AIR FORCES

	4 AF	2 AF
Number of crews	[illegible]	[illegible]
Number of individuals	[illegible]	2,[illegible]
Percent of crews changed	[illegible]	[illegible]
Percent of individuals changed:		
Officers	[illegible]	6
Enlisted men	[illegible]	9
Total	[illegible]	8
Percent of individuals changed, by phase:		
Phase I	3	3
Phase II	5	[illegible]
Phase III	5	[illegible]
Phase IV	3	[illegible]

[1]Does not include changes due to hospitalization, medical grounding, air-sickness, death or transfer for special duty. These add 1.1 percent to total men changed in class of 10 February and 5 percent to total men changed in class of 10 January 1945.

[2]Second Air Force training had only three phases.

The reasons for the crew changes in the 5 CCTS (H) classes in the Second Air Force are given in table 11.26. More than fifty percent of the changes resulted from medical disqualifications or incompetence of one of the members of the crew. Only a small percentage of changes was reported as due to incompatibility. These data are somewhat misleading since it was frequently reported, both by training personnel and by crew members themselves, that the real reasons for crew changes were not those specified in the official records. Because of the numerous difficulties involved in obtaining the real reasons for crew changes no further analysis was attempted.

TABLE 11.26.—*Reasons for crew changes in operational training*

SECOND AIR FORCE

Reasons	Number of changes			
	Officer	Enlisted	All	Percent
Personality clashes	2	11	13	7
"Buddy" deals, special friends	4	7	11	[illegible]
Emergency leave, AWOL	0	4	4	2
Medical, hospital, disqualified	28	51	79	[illegible]
Fear, refusal to fly	6	8	14	[illegible]
Incompetence, failure on job	[illegible]	24	27	13
Needs additional training, held over	3	16	[illegible]	8
Became instructors, special duties	3	3	6	3
Crew incomplete at start of training	[illegible]	4	6	3
Unknown reasons, miscellaneous	[illegible]	17	23	11
Total	[illegible]	145	202	100

The frequencies with which 70 crew members in the classes studied in the Fourth Air Force cited certain factors as important in causing crew [illegible] are given in table 11.27. From this table it is app[illegible]

sonality factors of one sort or another were regarded by crew members as the most frequent cause of changing personnel on crews.

Crew Formation in Terms of Compatibility

Originally, when the problem of crew matching was first considered, it appeared that for the most part no systematic procedures were in operation in the Third Air Force to match combat crew members so as to attain compatibility. A more systematic method was, therefore, devised by research personnel in cooperation with other personnel in the Air Force. As it was set up, the fundamental basis for crew formation was the elimination of extreme differences in various characteristics among the individuals forming a crew. The following factors were selected for matching of both officers and enlisted men in B–17 crews: rank, age, education and geograph-

TABLE 11.27—*Frequency of report as a factor in crew changes*

70 Combat Crew Members—Muroc, Fourth Air Force

Factor:	*Frequency of report*
Personality Conflict:	
Lack of cooperativeness and sociability	7
Wide age differences	7
Personality conflict cause unassigned	7
Religious and racial differences	4
Differences in marital status	4
Differences in regional origin	3
Differences in educational achievement	3
Differences in social background	2
Differences in rank	1
Hostility between combat returnees and new trainees	1
Poor discipline	6
Incompetence	6
Indifference to flying	5

ical location of the man's home. An additional criterion for officers was job proficiency and for enlisted men, sports interest.

Data on the above criteria, with the exception of the officer job proficiency, were obtained either from the WD AGO Form 66-2, or Form 20, or during the interview with the officer or EM on the processing line at the Third Air Force Replacement Depot, Plant Park. Several measures of job proficiency were obtained for the officers from the following records: AAF Training Command Proficiency Card (AFTRC Form 2); stanine scores recorded on WD AGO Form 64; and bombing accuracy (bombardiers only). Because of the nature of the criteria used and the availability of the records, EM and officers were first matched separately into partial crews and these partial crews were then matched to form complete crews. Insofar as possible, requests to be placed on the same crew with another individual were granted. Using the criteria mentioned above, 504

B–17 combat crews were matched at the Personnel Depot, Plant Park, just preceding their shipment to CCTS training stations.

Evaluation of Crew Assembly in the Third Air Force

After research personnel had formed B–17 crews on the basis of compatibility several attempts were made to evaluate the procedure used. Two promising checks of the results, the effect of the matching upon the number of crew changes during CCTS training attributable to noncompatibility, and ratings of crew compatibility by supervising officers were abandoned because no data were obtainable. Two further evaluations were made. First of all, the airplane commanders of 45 crews were interviewed by research officers who then rated the crews as being compatible or not compatible on the basis of the report of the airplane commanders. Independently, other research personnel made a rating as to the quality of the original matching. Of the 45 airplane commanders interviewed, 32 came from crews originally considered well matched and 13 from crews considered poorly matched. Chance agreement between the original matching of the crews and the ratings of compatibility would be expected in 56 percent of the cases. Table 11.28 reports the amount of agreement obtained between the original matching and judgments from the interviews as compared with what might be expected by chance. It will be noted that considerably more

TABLE 11.28.—*Compatibility ratings of well and poorly matched crews*

B–17 CREWS—THIRD AIR FORCE

Type of crew	Number	Rated compatible		Rated not compatible	
		Number	*Percent*	*Number*	*Percent*
Well-matched crews	30	25	83	5	17
Poorly matched crews	15	7	47	8	53
Total number	45	32	......	13	......

Agreement between matching and rating .... 33 out of 45 crews.
Agreement expected by chance .... 25 out of 45 crews.
Improvement over chance .... 12 (significant at 5 percent).

agreement was obtained than would be expected by chance alone. For the total group the agreement was complete for 73.3 percent of the crews. The difference between this and chance expectancy has a critical ratio of 2.2 which is significant at between the 1 and 5 percent levels.

A questionnaire was administered to combat crew members to obtain their opinions relative to the use of such factors as geographic location, education and age as matching principles in the formation of crews. In table 11.29 are presented the summaries of the answers to the questionnaires. Except for geographic location, a substantial majority of combat crew trainees supported the use of the factors used in the Third Air Force for the formation of crews.

TABLE 11.29.—*Proportion of crew members favoring consideration of certain factors in crew formation*

Factor	Percent favoring for officer crew	Percent favoring for enlisted crew
Officers:		
Geographic location	30	46
Education	90	88
Age	85	80
Enlisted Men:		
Geographic location	56	53
Education	76	72
Age	76	85

SUMMARY

With the continual development and revision of bombing techniques and the introduction of new equipment, the lead crew assumed an increasingly important role as the war progressed. Eight months after research psychologists were assigned to Continental Air Forces, procedures were established for the assembly of all crews by the Training Command and the designation of certain of these crews as potential lead crews. This designation was to be reviewed and changed if necessary on the basis of performance in operational training. Such a procedure necessitated thorough study of methods for evaluating the proficiency of crews. Soon after they were started by research personnel in the Second, Third and Fourth Air Forces, these studies were interrupted by the cessation of hostilities and the resulting closing down of training. Hence the data presented permit only tentative conclusions.

In the area of criteria, measures of bombing accuracy, especially those of radar bombing of industrial targets seemed particularly good as criteria of the proficiency of the crew since they were both moderately reliable and represented the proficiency of all members of the bombing team. Three rating scales were developed and analyzed; two in the Second and one in the Fourth Air Force. They were used in connection with B–29 training and found to be moderately reliable. One result that seems worthy of special comment in connection with the use of rating scales was the tendency for over-all judgments to be more reliable than ratings on specific items of behavior.

Rate of mission accomplishment and crew awards were subject to many criticisms and were not considered satisfactory as criterion measures.

The stanines of individual members of the crew showed no significant relationship to the criteria except for the stanines of the airplane commander which showed significant correlation with training accomplishment scores.

Composite stanines such as the average stanine of all crew members or the sum of their stanines showed almost no relationship to the various criteria of crew proficiency.

The evidence on the homogeneity of proficiency of crew members was

ambiguous. Rating scores of crew members showed moderate intercorrelations (probably due in part to "halo") while trainer and ground school grades showed zero correlation.

There was some evidence that the expressed interest of individual crew members in becoming lead crew members was related to instructors' ratings of them and of the crew as a whole as lead crew material.

Studies of crew changes in two air forces indicated that during operational training one or more changes in personnel took place on more than half of the crews.

Crew assembly techniques established in the Third Air Force with the aid of psychological personnel prior to AAF Letter 50-117 could not adequately be evaluated with the available facilities.

It is believed that answers to such fundamental questions as to whether the lead crew has characteristics above and apart from its members or is merely the sum of its constituent members will have to wait upon the development of more adequate criteria both of crew and individual proficiency.

CHAPTER TWELVE

Studies in the Acquisition and Retention of Aircrew Skills

INTRODUCTION

Since the major responsibility assigned to aviation psychologists in the Continental Air Forces was that of obtaining criterion data for test validation, little time was available for carrying out studies on the efficient acquisition and retention of aircrew skills. The few studies which were accomplished were necessarily limited in scope and opportunistic in regard to subject matter and design. In fact, the element of experimental design in the sense of differential group treatments and the systematic control of important conditions, can hardly be said to have obtained in any of the studies to be described in this chapter. Nevertheless, various analyses of the available training records have been made, some of which suggest important modifications of future operational training policies.

At the outset, the reader must acquire a broad understanding of the conditions under which measures of skill in operational training were obtained. Although many of the conditions have been discussed elsewhere in this volume, especially in sections dealing with the reliability of measures of proficiency, further discussion is appropriate here. It should be pointed out that filling the commitments for trained men to combat theaters enforced the strictest kind of adherence to time schedules imposed upon training bases by higher headquarters. This meant that conditions could seldom, if ever, be kept more than roughly comparable from one "trial" or practice period to the next. For example, it is known that numerous factors affect the ability of the fighter pilot student to secure hits on an aerial target. The airplane, the altitude, the direction of attack, the turbulence of the air, and the range and angle off at which firing is commenced are but a few of these factors. Since no systematic control of these factors was possible, it was only a chance matter if the pattern were the same on successive trials. As a matter of fact, from the standpoint of training for combat, constancy of conditions would have been highly undesirable. A similar problem existed in the case of bombardment skills. In synchronous and fixed-angle bombing the skill of the bombardier was affected by the airplane, the bombsight, the skill of the pilot, the turbulence of the air, the mechanical condition of the autopilot, and a host of other factors which varied from trial to trial and over which the bombardier had no control.

In radar bombing this complex of determinants reached a point where it was difficult for one to tell for sure just whose skill was being measured by a score such as circular error or to what extent the score was a function of personnel rather than of other factors. In the case of bombardment skills, available measures actually represent the collective proficiencies of persons working together as a team. Each member of the team makes some contribution to hitting the target, but the relative magnitude of the individual contributions is at present indeterminate. In the case of gunnery skills, the measures more nearly reflect the proficiency of single individuals. But the reader must view any curve of learning drawn to a sequence of these measures with the understanding that the conditions are not strictly comparable from measure to measure. However, in spite of factors such as these, it is believed that learning analyses are justified, if for no other reason than the fact that they describe operational training performance as it was actually found in the military setting.

In this chapter, evidence will be presented concerning (1) repetition of practice, distribution of practice, and transfer of training as factors affecting the acquisition of aircrew skills, and (2) the retention of fixed gunnery and bombing skills over periods of time in which they were not used on operational training missions.

FACTORS AFFECTING THE ACQUISITION OF SKILLS

Practice

In spite of the fact that most measures of aircrew skills have only moderate reliability, there is some evidence to indicate that such skills as fixed aerial gunnery, flexible aerial gunnery, and radar bombardment improve in a measurable way with practice.

Fixed Aerial Gunnery

Early studies, in which the percents of hits obtained on an aerial target were plotted against successive scorable missions, were somewhat ambiguous with respect to the demonstration of learning. From the data presented in table 12.1 it would appear that some increment in score was present in four out of the five classes represented.

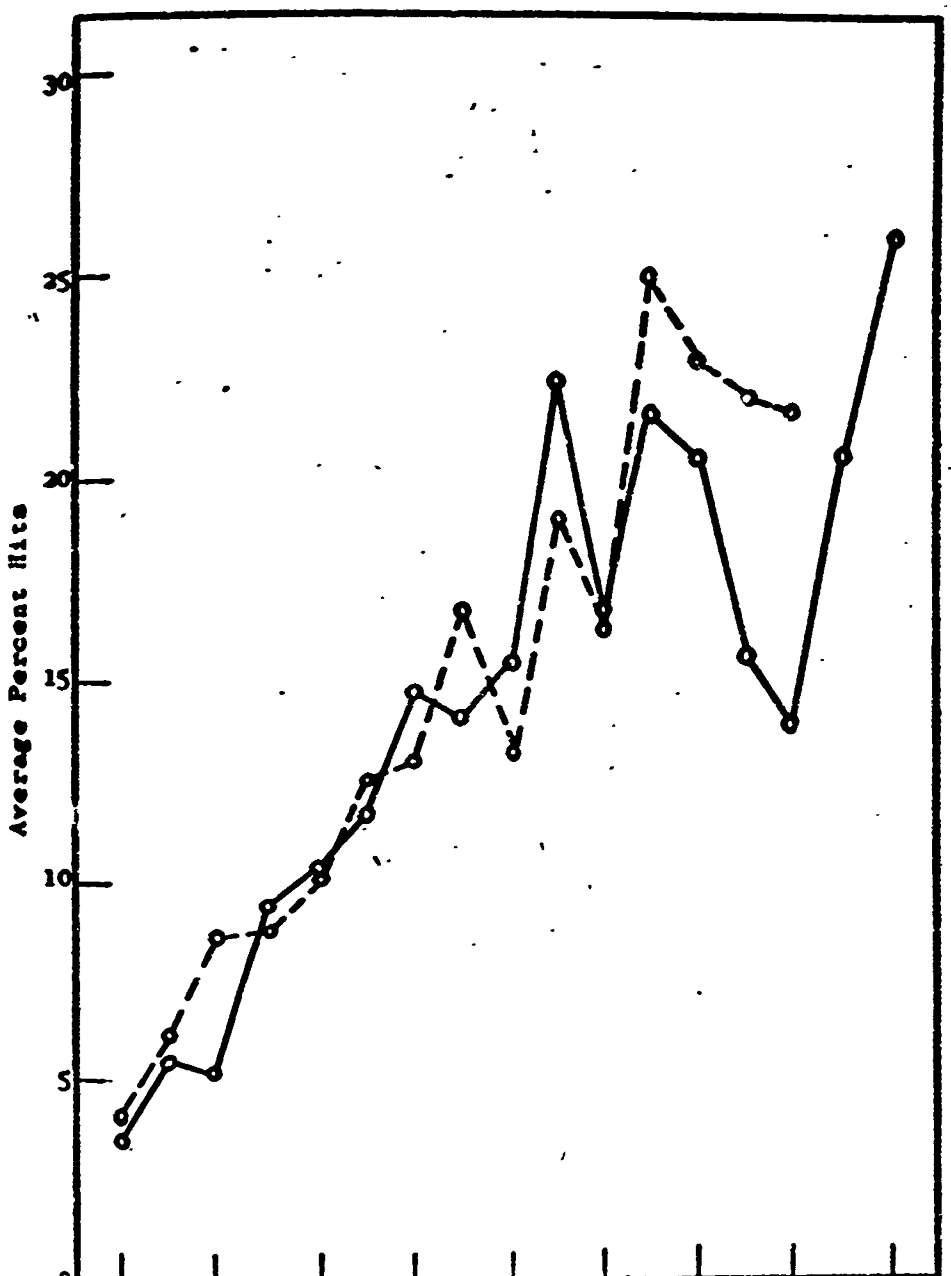
More conclusive results were obtained in later studies. In one of these it was possible to examine the scores of 2 small groups of instructors, some of

TABLE 12.1.—*Average percents of hits obtained on successive missions by five classes*
FIRST AIR FORCE, AUGUST TO OCTOBER

School	Mean percent hits on successive missions								
	N	1	2	3	4	5	6	7	8
Millville	108	2.7	1.6	2.8	2.1	2.6	2.3	2.3	1.9
Norfolk	79	3.6	4.6	5.3	4.9	5.4	4.8	.....	...
Bluethenthal	85	2.4	3.4	3.7	3.3	5.0	.....	.....	...
Bradley	87	2.0	2.1	3.2	3.6	.....	.....	.....	...
Dover	90	1.8	2.2	1.8	2.2	2.2	3.4	.....	...

whom had preserved their records for as many as 85 missions. When the percents of hits for successive groups of five missions were averaged and plotted against a time axis, the curves shown in figure 12.1 were obtained. These curves demonstrate small but definite increments per unit of practice, and, for all practical purposes, seem to be without acceleration. These data are somewhat ambiguous, however, since the population decreased from one group of five missions to the next. The data from which the curves

Figure 12.1—Increase in shooting accuracy shown by two groups of instructors after extended practice



were plotted are shown in table 12.2 in comparison with the results of student pilots whose records extended over not more than 15 missions.

Inasmuch as there were rather large individual differences in the shooting accuracy of fighter students at the outset, it was of interest to observe whether continued practice in gunnery effected a reduction of these differences. For this purpose one small group of instructor pilots was divided into high-scoring and low-scoring subgroups upon the basis of the accuracy demonstrated on the first 10 missions. The nine low-scoring instructors had a percent hit average of 3.2 while the nine high scorers had an average of 6.8. The learning curves of these groups, showing the average percents

Table 12.2.—Average percents of hits obtained on successive groups of five missions by instructors and student pilots.

FIRST AIR FORCE

Missions	Instructors percent		Instructors percent		Students percent	
	N	Hits	N	Hits	N	Hits
1–5	14	3.5	19	4.2	21	7.0
6–10	14	5.3	19	6.0	21	9.0
11–15	14	5.2	19	8.6	21	9.6
16–20	14	9.4	19	9.0	...	...
21–25	14	10.5	19	10.2	...	...
26–30	14	11.6	17	12.5	...	...
31–35	13	14.8	17	13.2	...	...
36–40	11	14.0	15	16.8	...	...
41–45	9	15.3	13	13.4	...	...
46–50	8	22.3	11	19.4	...	...
51–55	7	16.7	9	16.2	...	...
56–60	6	21.3	8	25.0	...	...
61–65	5	20.5	7	23.0	...	...
66–70	4	15.5	5	22.3	...	...
71–75	4	12.9	4	22.0	...	...
76–80	4	20.5	...	...	...	...
81–85	4	26.0	...	...	...	...

of hits for successive performance units of 5 missions each, are shown in figure 12.2. From this figure it is apparent that the advantage of the initially superior group is maintained for approximately 45 missions. After 45 missions the curves are extremely irregular, which is doubtless a function of the small number of cases.

The foregoing results would seem to indicate that learning fixed aerial gunnery is demonstratable during operational training. In fact, the data obtained from instructors show that there can be steady improvement over a long period of time beyond the requirements of operational training. There is no evidence that the upper limit of shooting accuracy had been reached within any of the amounts of practice studied.

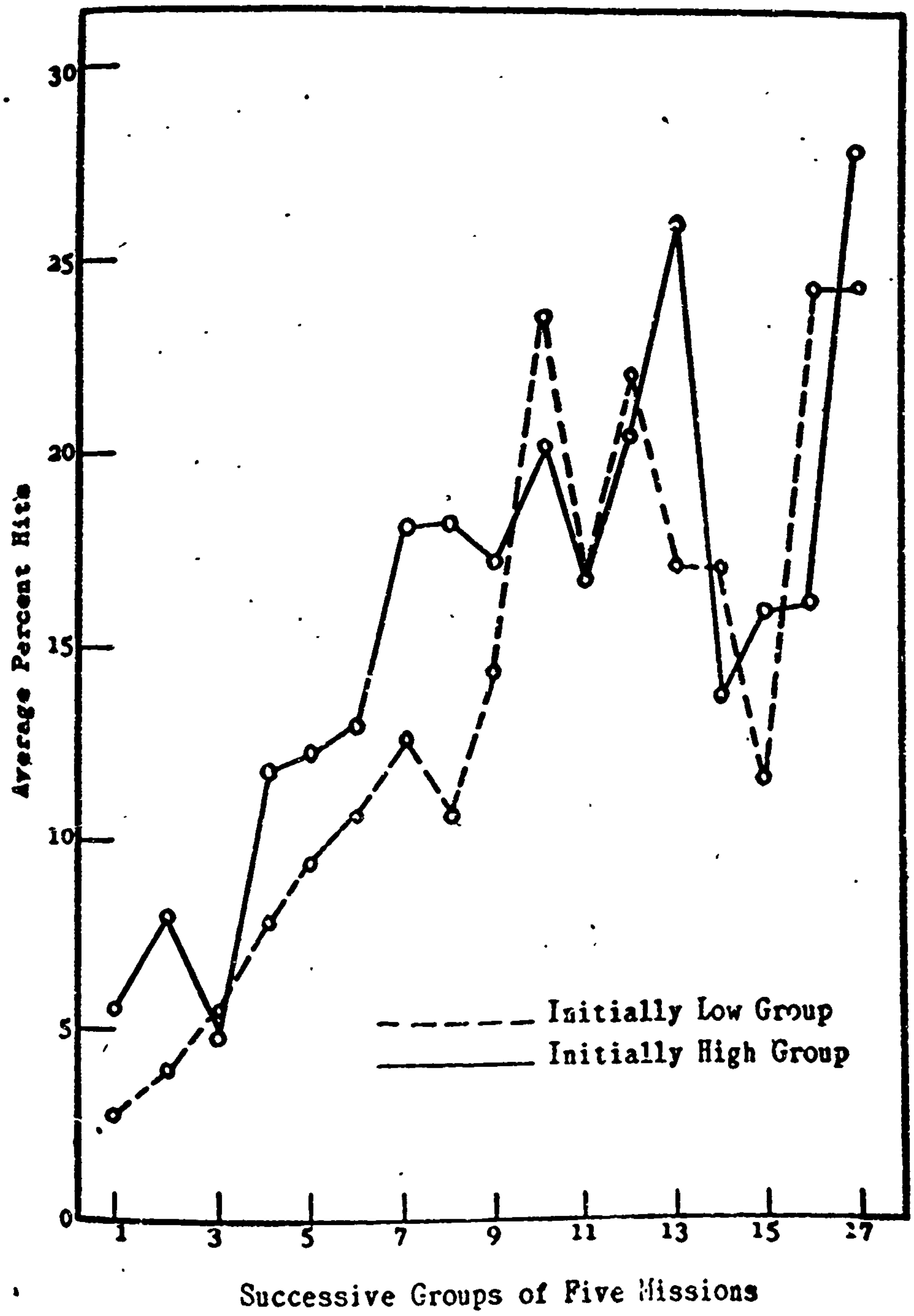
Flexible Aerial Gunnery

For the demonstration of learning in flexible aerial gunnery, small groups of data were available from a gunnery instructors school and from combat crew training station classes in the Second Air Force. These data, obtained from two successive gun camera missions, illustrate the progress to be expected with the use of both noncomputing and B–29 computing sights. The score representing proficiency with the computing sight is a composite

error score involving mil error, visual target size in percent, and an over-all judgment of the quality of the film. The presence of learning would be shown by a decrease in this score with practice. The score based upon the noncomputing sight (waist guns) is equivalent to a percent-of-hits score, with due allowances made for error in film assessment. This score would be expected to increase with practice if learning were present.

The changes in scores from the first to the second missions are shown in table 12.3. Changes in scores from the first to the second attacks in the

FIGURE 12.2—*Reduction of individual differences with practice*



first mission are shown in table 12.4. These data suggest that some increment in proficiency has been gained although because of the small numbers of cases the differences between the means is significant only in the case of the instructors using the computing sight.

TABLE 12.3.—*Changes in camera gunnery scores from the first to the second missions*

THREE SECOND AIR FORCE SAMPLES—B-29 GUNNERS

Source of data	First mission		Second mission		N
	Mean	S. D.	Mean	S. D	
Gunnery school instructors computing sights	23.30	13.10	16.70	6.79	30
CCTS class computing sights	1.66	.59	1.38	.58	27
CCTS class waist guns	35.80	18.80	41.70	17.60	[illegible]

TABLE 12.4.—*Changes in camera gunnery scores from the first to the second attacks in the first mission*

TWO SECOND AIR FORCE SAMPLES—B-29 GUNNERS

Source of data	First attack		Second attack		N
	Mean	S. D.	Mean	S. D.	
CCTS class computing sights	2.02	0.86	1.97	0.85	129
CCTS class waist guns	34.70	17.29	37.10	18.70	112

Navigation

Some slight evidence of learning in navigation was found in the ratings given to 430 navigators on three successive missions in the Second Air Force. The missions were graded on a three-point scale, with a "3" representing the best performance. The average ratings were 1.91, 2.05 and 2.13 with standard deviations of 0.66, 0.68 and 0.62 for the first, second and third missions respectively. The difference between the average rating for the third mission and that for the first mission is significant at the one percent level. There is some likelihood, however, that this increment in rating reflects nothing more than the instructors' beliefs that students should improve with practice.

Distribution of Practice

Training policies established by command headquarters usually specified the number of missions to be flown by fighter pilot students for the accomplishment of minimum aerial gunnery requirements. In the Fourth Air Force this number was originally set at six missions. In actual practice, however, the number of missions employed was found to vary from four to eight. For a sample of 221 students trained at Van Nuys and Ontario during the autumn and winter months of 1944, the distribution of missions flown to accomplish 2,000 rounds of aerial gunnery is shown in table 12.5. From this table it is evident that some students used twice the number of missions to meet minimum gunnery requirements as did others. It was thought possible that the factor of number of missions would bear some consistent relationship to final proficiency.

In order to investigate the possibility that the number of missions had some consistent relationship to final proficiency, the individual aerial gunnery mission records of 3 classes (Fourth Air Force, Ontario A and B, Van Nuys III) were examined. For each student the mission upon which he fired his 2,000th round was noted, together with his percent of hits for that mission. Discarded from the analysis were cases in which the target

TABLE 12.5.—*Number of fixed gunnery missions flown to accomplish minimum requirement of 2,000 rounds*

221 P-38 TRAINEES—FOURTH AIR FORCE

Number of missions	Percent of pilots
8	13
7	14
6	38
5	27
4	8
Total	100

had been lost or shot off on the final mission. The remaining cases were then sorted into mission-frequency groups within classes. The average percent of hits was computed for each mission-frequency group. Table 12.6 presents these data for each of the classes.

From table 12.6 it is evident that terminal aerial gunnery proficiency is a positive function of the number of missions over which the required rounds are distributed. It is believed that this finding is in accordance with results usually obtained in distributed practice experiments, although

TABLE 12.6.—*Average percent hits on final aerial gunnery mission in relation to number of missions flown to accomplish minimum of 2,000 rounds*

221 P-38 TRAINEES—FOURTH AIR FORCE

Missions	Ontario A		Ontario B		Van Nuys	
	N	Percent	N	Percent	N	Percent
8	13	3.3	0	....	15	7.1
7	13	1.9	3	3.2	15	5.6
6	37	1.9	27	2.6	21	5.5
5	33	1.5	19	1.6	7	4.7
4	12	0.9	6	1.1	0	......

it should be noted that the temporal arrangement of "practice" and "rest" elements is somewhat atypical. In the typical distributed practice experiment either the length of time between practice periods or the amount of practice in a single continuous session are systematically varied while other factors are controlled. In aerial gunnery, however, the design is not so clear-cut. Obvious enough is the fact that 2,000 rounds distributed over 8 missions will tend to be fired at a lower rate per mission than they would be if distributed over only 4 missions. If it is assumed that the number of rounds fired per pass at the target is fairly constant—and instructors report that this is the case, though they may be wrong—it follows that the students who required eight missions have made fewer passes per mission than

those who required only four missions. This state of affairs roughly approximates that of the typical distributed practice experiment, since the passes (units of work) are more widely separated in time for the eight-mission students than for the four-mission students.

Of course, it is possible that other factors, not necessarily related to the fact of separation of work-units in time, may provide at least partial explanation of these results. For example:

(1) More frequent missions provide more numerous opportunities for both student and instructor to check upon the accuracy of gunnery, i.e., there is fuller knowledge of results.

(2) More frequent opportunities are available for the recognition and correction of errors in handling the aircraft.

(3) Students who require more missions may have been basically more proficient pilots. This would imply that their mission frequencies reflected a meticulous economy of ammunition, i.e., rounds were fired with deliberation and intention to score. This seems a logical possibility even though instructors generally reported the number of rounds per attack were relatively constant.

In any case, the results are more suggestive than conclusive. In view of the military importance of high proficiency in aerial gunnery, the problem of the distribution of practice should be investigated thoroughly and systematically.

Transfer of Training

The factor of positive transfer holds considerable significance for aircrew training, for upon the extent to which it can be realized depends the successful use of synthetic training devices. It is obvious that if transfer can be effected from performance in mock-ups to performance in the air, a considerable saving can be made in materials to say nothing of time necessarily expended in training.

Upon the *a priori* assumption that transfer of training could be obtained, the Army Air Forces has made extensive use of synethetic training devices. One of these was the Gunairstructor. The Gunairstructor is a training device for fixed gunnery, intended to simulate aerial combat and to provide practice in control coordination, aircraft alignment, range estimation, and deflection allowance. The equipment was mounted in a specially constructed dark room. The student, sitting in a standard fighter aircraft cockpit equipped with regulation controls, fired at the image of a moving target (an airplane silhouette) controlled by the operator. The image was projected upon a screen in front of the student's cockpit. In order to register hits the student had to establish the correct deflection allowance and alignment. When hits were scored the target blinked red and an automatic device recorded the number of rounds fired and the number of hits scored. Since the path and attitude of the target airplane were controlled by the operator, it was possible for him to present problems which varied

in difficulty. Individual training sessions were usually 15 minutes in length.

In view of the emphasis placed on the Gunairstructor as a training device, it became important to determine whether its use resulted in an increase in aerial gunnery proficiency. Although the exigencies of training prevented the execution of an experiment designed to evaluate the efficacy of the device under optimal conditions of use, it was nevertheless possible to examine available data for the purpose of determining, under actual conditions of use, (1) whether amount of practice on the device was associated with aerial gunnery proficiency (2) whether gunnery proficiency and Gunairstructor proficiency resulted from a common set of determinants, and, if not, (3) what conditions of actual operation produced the presumed communality of the two performances.

Amount of training.—In order to ascertain the effect of varying amounts of training upon aerial gunnery proficiency, the data from two fighter pilot classes (Fourth Air Force, Van Nuys I and II) were examined. For each class the number of training sessions engaged in by the students varied from 4 to 25. The number of sessions was administratively determined and was not a function of the ability of the student as far as could be determined. The product-moment correlation between size of aerial gunnery score and number of synthetic training sessions was, therefore, directly interpretable. These correlations, with distribution statistics, are shown in table 12.7. It is evident from these results that the aerial gunnery performance of fighter students with many synthetic training sessions was not significantly different from that of students with few such sessions.

TABLE 12.7.—*Correlations between the number of Gunairstructor training sessions and percent hits in aerial gunnery*

FOURTH AIR FORCE

Class	N	Sessions, 1		Gunnery, 2		r_{12}
		M_1	SD_1	M_2	SD_2	
Van Nuys I	117	15.10	4.30	4.02	1.74	−0.09
Van Nuys II	132	13.61	3.06	2.37	1.47	.08

Communality of aerial and synthetic performances.—If the Gunairstructor has rationale as a synthetic training device for fixed gunnery, it would seem that the two performances should be so similar as to be significantly correlated. The student who scores many hits in the Gunairstructor does so presumably because he has learned to solve the problems of range, deflection, alignment and control coordination. Solving these problems on the ground should bear some relation to solving them in the air. In order to determine the extent of this relationship, the product-moment correlation between percent hits in the training device and percent hits in the air was computed for one class (Fourth Air Force, Van Nuys II). The results, shown in table 12.8 indicate that the two scores exhibit no communality.

Factors reducing the efficacy of the Gunairstructor.—In an effort to explain the foregoing results, additional studies of synthetic training were

performed. In these studies attention was directed to the operators, since it was observed that they were permitted considerable freedom in such matters as problem difficulty and orderliness of instructional sequences. In addition, they were required to make periodic ratings of their students' progress with a view toward correction of individual deficiencies. A lack of system and/or objectivity in either of these operator functions might

TABLE 12.8.—*Correlations between percent hits obtained in the gunairstructor and percent hits obtained in aerial gunnery*

FOURTH AIR FORCE

Correlation	N	Mean, 1	SD, 1	Mean, 2	SD, 2	r_b
Sum of percent hits in gunairstructor, sessions 1–[illegible] (1) x percent hits in aerial gunnery (2)	116	198.97	26.25	2.46	1.49	−0.[illegible]
Sum of percent hits in gunairstructor sessions 11–16 (1) x percent hits in aerial gunnery (2)	116	208.12	27.20	2.46	1.49	.[illegible]

largely account for the results of the foregoing inquiries into the efficacy of the training device. Accordingly, an analysis was first made of the differences among operators in the manner of instruction. The data (percent hits) for a single practice session for 312 students in 3 classes (Fourth Air Force, Van Nuys I, II, and III) were sorted by class and then by operator within each class. The variances attributable to classes, operators, subclass discrepance and error were determined. As shown in table 12.9 the significance of the first three variances was established by the F-test.

TABLE 12.9.—*Analysis of variance of hits scored in the gunairstructor on a single practice session*

VAN NUYS I, II, AND III, FOURTH AIR FORCE

Source of variance	Sum of squares	df	Mean square
Classes	3,388	2	**1,6[illegible]
Operators	3,112	4	**778
Subclass discrepance	2,742	8	**3[illegible]
Error	28,511	297	96
Total	37,753	311	....

These results indicate (1) that there were significant differences in scores between class groups, (2) that instructors differed from one another in the scores which they gave and (3) that a given instructor's scoring tendency varied from class to class.

The adequacy of synthetic instruction was examined further by an analysis of the skill with which operators could diagnose and correct individual deficiencies in lead estimation, angle estimation, and control coordination. It was assumed that, under adequate conditions of instruction and rating of performance, the scores made on the Gunairstructor would exhibit some communality with ratings by instructors. The analysis consisted of the computation of biserial coefficients of correlation between ratings and scores for each of three items rated. The discrete variable consisted of a below-average/above-average dichotomy. The three biserial coefficients thus obtained ranged from 0.06 to 0.19. If it is granted that the scoring

of hits is the only valid measure of the quality of performance, then it is evident that operators were unable to distinguish between good and bad performance. It follows that they could not reasonably have been expected to diagnose and correct individual deficiencies, with the result that a general transfer of training could not have been brought about thereby.

What has been demonstrated in the above analysis is that the profit to be gained from training on the Gunairstructor under haphazard conditions of instruction is negligible. Crucial studies involving the careful control of training conditions must be made before a final evaluation of the Gunairstructor, or any other synthetic device, can be accomplished.

THE RETENTION OF SKILLS DURING PERIODS OF DISUSE

In the handling of assignments and shipments of large numbers of men, it is inevitable that some individuals will be subjected to periods of inactivity as regards the practice of their special skills. It is of great practical importance, therefore, to determine whether these periods of disuse of skills result in a reduction of proficiency.

Two studies were undertaken in the Fourth Air Force to determine the extent to which varying periods of lack of practice affected proficiency. In one study of 300 fighter pilots who completed P-38 training, it was found that the average air-to-ground gunnery score of those who were inactive for as long as 200 days was only 6 percent poorer than that of pilots who were inactive for not over 50 days. In the other study, involving bombardiers, it was found that the average circular bombing error for those who were inactive for as long as 90 days was only one percent poorer than that of bombardiers who were inactive for no more than 30 days. For the periods of time studied, therefore, the loss of proficiency resulting from periods of inactivity appears to be negligible.

SUMMARY

Several studies of the acquisition and retention of aircrew skills have been reviewed. In these reviews the reader's attention has been called to the difficulty of performing adequate and crucial investigations under the exigencies of wartime training. In spite of these difficulties, however, certain problems have been studied with some profit. Learning in several skills was demonstrated. Lack of systematic control of training conditions was shown to prevent the clear demonstration of transfer of skills from synthetic settings to aerial situations. Finally, some light was thrown upon the question of what happens to aircrew skills during periods of inactivity. There are, of course, many other problems in this area which warrant investigation. If useful results are to be obtained, however, these studies must be performed with some regard for the principles of experimental design.

CHAPTER THIRTEEN

Attitudes of Aircrew Personnel Returned From Combat Toward Further Duty

INTRODUCTION

At the direction of the Research Division, Office of the Air Surgeon, Headquarters Army Air Forces, the Research Section of the Third Air Force conducted a survey of the attitudes of aircrew officers and gunners returned to the United States after a tour of combat duty. The survey was concerned chiefly with the attitudes of these returnees toward their further utilization in the Air Forces, including their attitude toward a second tour of combat duty. An analysis was made of the effect of factors such as age, grade, and severity of combat experience on the attitude of these returnees. Information obtained from questionnaires was supplemented by interviews with returnees, commanding officers, training officers and flight surgeons. Since the survey of the attitudes of returned aircrew *officers* and the study of returned *gunners* were carried out independently, the two studies are discussed separately on the following pages.

SURVEY OF RETURNED AIRCREW OFFICERS

Method

Research personnel visited 26 out of 38 main stations in the Third Air Force administering a questionnaire to a representative group of 528 returned aircrew officers in all. This sample was carefully selected to resemble as closely as possible the total group of returned aircrew officers in the Third Air Force with respect to grade, rating, (pilot, bombardier, navigator), combat theater and length of time in the United States since return from combat. When time permitted, information obtained from the questionnaire was supplemented by interviews with the returnees.

The questionnaire (*see* appendix I 1) submitted to the officers consisted of 77 items covering the following areas:

Willingness to return for second tour of combat duty. Length of stay in the United States.

Combat experiences.

Attitude toword further training.

Present job assignment and job satisfaction.

The responses of the returnees were recorded on IBM punch cards. Sample sortings were made on all items. Interitem counts were made to determine the effect of various factors upon the returnees' attitudes as reflected in certain critical items. The difference between groups which are mentioned in the following discussion are statistically significant at least at the 5 percent level. In some comparisons, responses to similar items were pooled in order to increase the number of cases upon which to compute the significance of the difference. The sample questionnaire in appendix I.1 includes a tabulation of the percentage of responses to each alternative of all items.

Results

General Characteristics of Returned Officers

An accurate interpretation of the findings of this survey requires a knowledge of the composition of the sample. The following paragraphs summarize pertinent characteristics of the returnees.

1. *Personal characteristics.* a. *Grade.*—Slightly more than half of the returnees were first lieutenants; one third of them were captains.

b. *Age.*—About half of the returnees were between the ages of 24 and 26.

c. *Marital status.*—Slightly more than one-third were single. Of the 61 percent who were married, 43 percent had married since their return from combat.

d. *Dependents.*—Slightly more than one-third had no dependents; 61 percent had 1 or 2 dependents, including wife.

e. *Education.*—Twenty-eight percent had no college training; 52 percent had from 1 to 3 years college, and the remaining 20 percent had 4 years or more of college.

f. *Time in the United States since return from combat.*—More than two-thirds of the returnees had been back from 3 to 12 months.

2. *Facts about combat tour.*—a. Approximately 60 percent were pilots or copilots, 19 percent were navigators and 21 percent were bombardiers.

b. About 20 percent flew in group lead positions as pilots, bombardiers or navigators while in combat.

c. Theaters were represented in the samples as follows: European 17 percent, Mediterranean 44 percent, Pacific 27 percent, others 12 percent.

d. Fifty-seven percent of the group had been overseas from 9 to 14 months inclusive.

e. The median number of missions was in the interval between 41 to 50. Seventeen percent reported 70 or more missions.

f. Slightly less than half of the group flew between 100 and 200 combat hours; 30 percent flew between 200 and 300 hours.

g. Ten percent completed less than the required number of missions for their theaters. Forty-one percent reported no policy in effect relative to the required number of missions.

h. More than half of the sample were assigned , medium and light

bombers, 22 percent flew heavy bombers; 19 percent flew fighter aircraft.

i. Rotation policy was responsible for the return of 55 percent, 28 percent were returned because of combat fatigue, wounds or injuries.

j. Only 5 percent of the sample were on limited service status at the time of study.

3. *Severity of combat experience.*—a. Almost half of the returnees had experienced forced landings in friendly territory. Seven percent were forced down in enemy territory.

b. More than two-thirds met enemy opposition on every or nearly every mission.

c. Six percent had bailed out in either friendly or enemy territory.

d. Nineteen percent were wounded in combat; 12 percent sustained injuries in connection with aircraft in non-combat situations.

e. Twenty-nine percent reported having suffered from operational fatigue.

f. Thirteen percent reported having members of their crew killed, 42 percent reported wounds among their crew members.

g. Almost half (40 percent) reported a fourth or more of their close friends missing, killed or wounded.

h. Eight percent reported experiencing air-sickness sometime during their combat tour.

4. *Current duty assignment.*—Sixty-five percent of the officers were assigned as flying instructors; 17 percent had flying duties other than instructing. Five percent were ground school instructors and 13 percent were assigned to administrative duties.

Willingness to Return to Combat

Three questions were so worded as to determine in a specific way the willingness of these aircrew officers to return to combat. The first question was concerned with the attitude of the officer toward return within various periods of time; the second was concerned with the willingness to volunteer for a second tour; the third was concerned with the strength of desire to return to combat. Thirteen percent indicated they would return at once. An additional 15 percent would return within 6 months. More than half of the officers would not volunteer for a second tour and only 30 percent expressed any desire at all to return to combat. The three questions were found to be highly reliable in that if an individual answered one question by indicating great willingness to return, he answered the other questions with responses indicative of great willingness. The same was true of those indicating a reluctance to return; their responses on the other questions indicated a similar reluctance.

Factors Affecting Willingness to Return to Combat

The willingness of officers to return was found to be related in various ways to other variables on which data were collected. These factors are listed below.

1. *Rank.*—There was no clear cut evidence that rank was related to willingness to return for another tour, although it was found that captains were more willing to return than were first lieutenants.

2. *Marital status.*—Married officers, regardless of whether they were married before or since their tour of duty, were less desirous of returning to combat than were single officers. This was verified by the fact that family ties and obligations were given most often as the reason for not desiring to return to combat.

3. *Length of time in United States.*—The longer officers had been in this country the more willing they were to return to combat duty.

4. *Duty assignment.*—In general, officers assigned to flying duty were more willing to return than officers in other assignments. A greater percentage of officers who were assigned as operations officers were willing to return than of officers in any other military occupational specialty.

5. *Military status.*—A larger percentage of officers who were in the Regular Army or National Guard were willing to return than officers who had no such connection.

6. *Amount of combat experience.*—In general, the greater the number of missions completed the less willing the officer to return to combat.

7. *Type of aircraft used in combat.*—A greater percentage of those officers who had combat experience in B–26 type aircraft expressed willingness to return than was the case with any other aircraft. Fighter pilots showed the next highest percentage willing to return.

8. *Severity of combat experience.*—The more severe the enemy opposition encountered the greater the proportion of individuals who expressed a strong desire not to return to combat. A greater percentage of officers who had histories of combat fatigue were unwilling to return than of officers without such histories.

9. *Degree of satisfaction with general conditions in the United States.*—More of the officers who expressed dissatisfaction with military courtesy, discipline and training in the Zone of Interior were willing to return than was the case for officers expressing dissatisfaction with other conditions.

10. *Attitude toward assignment.*—Those officers who were indifferent to their assignment were more willing to return than were officers who were either satisfied or dissatisfied with their present jobs. A greater proportion of officers who felt that little use was being made of their ability or experience were willing to return than of officers who felt good use was being made of their abilities. More of the officers who desired flying duties not involving instruction were willing to return than officers desiring other types of assignment. Officers who desired ground duty were least willing to return to combat.

11. *Attitude toward combat effectiveness.*—More of the officers who felt that their stay in the United States since returning from combat had increased their combat efficiency were willing to return to combat than officers who felt their efficiency had been unaffected or decreased.

12. *Desire for retention in the army.*—Officers who desired to remain on active duty with the Air Forces (Regular Army) were more willing to return than were officers who desired to return to civilian life.

13. *Future combat conditions.*—Sixty-two percent of the returnees indicated they would be more willing to return with a crew all of whose members had combat experience. Slightly more than half would be more willing to return if given an assignment not involving frequent combat flying. Assignment to very heavy bombardment aircraft would also increase willingness to return to combat.

Factors Not Related to Willingness to Return to Combat

No relationship was found between willingness to return for a second tour of combat duty and any of the following variables: Age, aeronautical rating, length of time overseas, number of combat flying hours, theater of previous combat experience and number of friends who were casualties.

Motivation During Combat

Although somewhat outside the scope of the survey, an attempt was made to determine the relative strength of various motivating factors during combat. The two factors checked by almost three-fourths of the returnees as having greatly influenced them while in combat were: Desire to do a good job and preference for flying duty rather than ground duty. Less than half of the returnees were greatly influenced by such factors as: The excitement of combat, not wanting to let other crew members down, being ordered to perform the duties by authorities or regarding their assignments as the best way to strike the enemy. Only a small percentage reported that they were strongly motivated to return to combat by such factors as desire to square accounts with the enemy because of the loss of friends, preference for combat duty over present assignment, dissatisfaction with conditions in the United States or a desire to be in the middle of things rather than on the sidelines.

Utilization of Returned Officers in the Zone of the Interior

1. *Present assignment.*—The duties to which the officers were assigned together with the percentages in each are shown in table 13.1.

2. *Satisfaction with job assignment.*—More than 70 percent were satisfied with their job assignment and felt that good use was being made of their ability and experience. Officers who held lead positions in combat

TABLE 13.1.—*Duty assignments and preferences of aircrew officers*

Duty	Percent assigned	Percent desiring assignment[1]
Flying duty as instructor	65	34
Flying duty not as instructor	17	51
Ground school instructor	5	8
Administrative ground duties	13	31

[1] Individuals were permitted to express preference for more than one assignment.

were less often satisfied with their duties than officers without that experience. Ground school instructors were less satisfied than officers in other assignments. Officers who had been in the United States longest were less satisfied than officers more recently returned.

3. *Preferred assignment.*—The returned officers expressed preferences for assignment as shown in table 13.1. Also shown are the duty assignments of these officers. There are marked discrepancies between the actual assignments made and the preferences regarding assignments. However, the fact that more than 70 percent expressed themselves as satisfied with their assignments suggests that a general good adjustment was being made by the returned officers.

Certain variables were found related to duty preferences. There was a general tendency for the proportion desiring administrative duties to increase with increasing rank. The proportion of officers who desired administrative duties increased with the amount of education they had received. A greater proportion of pilots desired noninstructional flying duties than did navigators and bombardiers. Officers who had considerable responsibility in combat were less desirous of flying instructor assignments than officers with less responsibility. More fighter pilots preferred flying instruction duties than bombardment pilots. Over 60 percent of the officers who were dissatisfied with their assignment preferred flying duties other than as an instructor. More of the officers who desired to remain in the Air Forces preferred flying duties than officers who had no desire to remain.

4. *Present combat efficiency.*—Slightly more than half of the officers felt that their stay in the United States had increased their efficiency for future combat duty.

5. *Attitude toward further training.*—Almost all (95 percent) of the returnees desired further technical training in the Army. Pilots expressed a desire for engineering training. Most of the navigators and bombardiers desired pilot training. Although given the opportunity to select training other than in Army specialties, the subjects most frequently selected fell in this category.

6. *Post-war army plans.*—More than half of the officers expressed a desire to remain on active duty with the Air Forces (Regular Army). Slightly more than one-third preferred Reserve Officer Status. Only 4 percent desired to return to civilian life without retaining any army affiliation. More of the officers with the rank of major and captain preferred Regular Army status than officers of lower rank. Married officers were more interested than single officers in such an assignment. More pilots were interested in a career in the Regular Army than other aircrew specialists.

Comments of Commanding Officers, Training Officers and Flight Surgeons

Station commanders, flight surgeons, directors of training, directors of administration, air inspectors and administrative inspectors were in sub-

stantial agreement that all of the returnees eventually adjust to their assignment in the training air force situation. On the average, this adjustment takes about 6 months. The most important problems of adjustment arise from restrictions due to army regulations and to the rigidity of flying and training schedules. These officials felt that most returnees could return to combat after about nine months. Only from five to ten percent of the returnees should be considered for change of occupational specialty and/or reassignment. These officers believed that a short, intensive course in army administration should be instituted for returnees inasmuch as most returnees had no administrative experience.

SURVEY OF RETURNED GUNNERS

Method

Research personnel interviewed 364 returned gunners and administered a questionnaire to 2,659 more. To supplement the information obtained from interview and questionnaire, 124 officers (commanding officers, flight surgeons, and training officers) were interviewed. The questionnaire (see appendix L2) consisted of 85 items and covered the same areas as did the officer questionnaire described above. The method of analyzing the results was the same as that discussed earlier for the returned officers. Although the gunners were asked to sign their questionnaire, it was explained that names were desired only to facilitate obtaining further information from their Soldier's Qualification Cards. It was felt that frank answers were obtained and that the results could be considered accurate reflections of the attitudes of the individuals sampled. The results of the interviews substantiate in all respects those obtained from the questionnaire and therefore are not presented separately in the following discussion. The sample questionnaire in the appendix gives the percentage of the gunners responding to each alternative of all the questions.

Results

General Characteristics of Returned Gunners

1. *Personal characteristics.* a. *Grade.*—More than 90 percent of the gunners were high ranking noncommissioned officers (staff and technical sergeants).

b. *Age.*—More than 60 percent were between the ages of 21 and 26 years of age.

c. *Marital status.*—Slightly more than half of the gunners were single, with no dependents.

d. *Education.*—About 60 percent had had 3 to 4 years of high school or vocational school. A very small percentage had training beyond high school.

e. *Time in the United States since return from combat.*—About two-thirds of the gunners had been in the United States from 1 to 7 months since return from combat.

2. *Facts about combat tour.*—a. Approximately one-fourth were mechanical specialists such as engineers, crew chiefs or mechanics in addition to being gunners while in combat; slightly more than one-third were armorer gunners; one-fifth were radio or radar gunners and slightly more than one-tenth were gunners only.

b. More than half flew all or most of their missions in B-17 type aircraft. B-24 gunners were next in frequency (22 percent) followed by B-25 gunners. A negligible number were assigned to other aircraft.

c. Thirty-six percent were based in England; 35 percent were from the Mediterranean Theater. The remainder were principally from the Pacific Theater.

d. The gunners were distributed fairly equally among the various gun positions except for the nose position which had markedly fewer representatives. The relatively small number of nose gunners is probably explained by the fact that bombardiers frequently fired the nose guns.

e. Most of the gunners had been overseas from 7 to 14 months and had flown from 20 to 50 missions.

f. About half of the gunners flew exactly the number of missions required by the theater. A fifth reported no policy in effect regarding number of missions.

3. *Severity of combat experience.*—a. Slightly less than half of the gunners had experienced crash landings.

b. More than 80 percent reported enemy opposition on all or almost all missions.

c. About 10 percent had bailed out of disabled aircraft.

d. Twenty percent were wounded in combat.

e. Almost half of the gunners were given some sort of treatment for operational fatigue.

f. About one-fourth of the gunners reported fatalities among their crew members; more than half reported wounds among fellow crewmen.

g. The majority of the gunners were returned as a result of the rotation policy prevailing at the time of their tour.

Willingness to Return to Combat

Only 6 percent of the gunners were willing to return to combat immediately. Sixteen percent had "some" desire to return or were indifferent toward return. The majority were not desirous of a second tour. More than half preferred being assigned to duty as basic soldiers to being returned to combat. Approximately half indicated they would be willing to return only after all available trained gunners had been sent to combat.

Factors Affecting Willingness to Return to Combat

1. *Marital status.*—Single men were more willing than married men to return for a second tour.

2. *Amount of combat experience.*—Gunners who had spent more than 18

months overseas during their first tour were more willing to return than gunners who had spent less time.

3. *Theater.*—Gunners from other theaters than the European, Mediterranean and Pacific were more willing to return than gunners from those theaters. Gunners from the China-Burma-India, Carribbean, Alaskan and other theaters were inclined to explain any lack of willingness to return to combat on the grounds that they had made their combat contribution. Gunners who had experience against the Germans, on the other hand, were reluctant to return because they felt that they could not stand the strain of further combat.

4. *Type of aircraft.*—In general, type of aircraft had no effect upon willingness to return.

5. *Crew position.*—Radio gunners were less willing to return than gunners in other positions.

6. *Severity of combat experience.*—Specific factors indicative of the severity of combat experience, such as number of bail-outs, crash landings, injuries or fatalities among fellow crewmen were not found related to willingness to return. However, the general factor related to severity of combat experience, namely, history of operational fatigue, did show such a relationship. Gunners with histories of operational fatigue were less willing to return than were gunners without that history.

Utilization of Returned Gunners in the Zone of Interior

1. *Job assignments.*—The gunners in the sample were assigned as follows: 57 percent instructors on flying status; 21 percent ground instructors in gunnery, armament, radio, mechanics, etc.; 10 percent ground duty not involving instruction; 3 percent instructing other than above; 1 percent clerical and administrative work; 8 percent no assignment.

2. *Job satisfaction.*—Sixty-eight percent of the gunners were satisfied with their job assignments. Seventy-two percent felt that the army was making good use of their ability and experience. The longer a gunner had been in the United States the more likely it was that he was satisfied with his assignment.

Attitude Toward Further Training

About 40 percent felt they needed no further gunnery training. Twenty-seven percent felt the need for more training in air-to-air firing. Certain important courses had been omitted in the original training of these gunners, as evidenced by the fact that 53 percent had had no gun camera missions and 29 percent had no course in position firing. With regard to additional training in specialties other than gunnery, 37 percent indicated they needed no further training. Additional training in such areas as armament, mechanics and radio operations was considered necessary by from 10–15 percent of the gunners. Fifty percent of the gunners reported that they were interested in attending school while in the Army, in prepara-

tion for return to civilian life, provided the courses were not restricted to army specialists. Courses most often requested were in mechanics, business, radio and radar, special trades and high school subjects.

Comments by Commanding Officers, Training Officers and Flight Surgeons

The officers interviewed were in substantial agreement that the majority of the gunners were making an adequate adjustment upon return to the United States. This adjustment, in their opinion, took from 3 to 6 months. The officers stated that surplus gunners could be absorbed by return to combat, reclassification to other jobs and by replacement of gunnery instructors without combat experience. Only about 10 percent were in definite need of immediate discharge, according to officers interviewed. The officers felt that the returned gunners were in need of further "on-the-job" training.

CHAPTER FOURTEEN

Summary and Evaluation

INTRODUCTION

In this final chapter, an attempt will be made to summarize the more important conclusions which came out of the work herein reported, and to evaluate the results obtained. In this connection the point of view from which judgments are made is an important consideration. In the case of the present report, at least three viewpoints can be taken, each with some reason. The material can be reviewed in terms of the formal missions of the psychological sections, as described in chapter 1, with conclusions stated in terms of how completely the stated objectives were attained. Again, the report may be examined to discover what contributions have been made to Army Air Forces selection and training procedures. This type of evaluation can be made only in terms of possible post-war contributions, since this research resulted in only minor changes in operating procedures in the Continental Air Forces during the war. From a third frame of reference, the results may be reviewed as a contribution by psychologists to the methods and subject matter of aviation psychology. As was the case with other sections of the Aviation Psychology Program, the research workers in the Continental Air Forces enjoyed an unusual opportunity for application of psychological principles to problems of selection and training involving very large populations, larger than are available in most peacetime military, educational, or industrial situations. This latter point of view would place chief emphasis on the methods used to secure the data, on the problems (and in some cases the anomalies) arising in the statistical handling of the data, and on the development of criteria for judging the practical usefulness of relationships discovered.

Throughout this chapter, the attempt will be made to discuss the research in terms of all three of these methods of evaluation. However, the writers have perhaps been most concerned with evaluation of the work in terms of a scientific project, as is indicated in the last of the three approaches. This position has been taken advisedly, in line with the general conviction that the largest national service can be rendered in a continuous effort to revise experimental hypotheses, sharpen research tools and perfect the methods used in investigating fundamental or underlying facts whether the phenomena be physical, psychological or social.

The evaluations offered will be confined to a discussion of the problems and findings presented in the chapters on the seven aircrew positions, on the crew as a whole, and on studies in the acquisition of skills. No further

remarks will be made on the study of attitudes of aircrew personnel returned from combat, done by the Third Air Force, nor on the miscellaneous studies reported in chapter 12. While these were undertaken under either broad or specific provisions of directives from Headquarters, Army Air Forces, they were somewhat aside from the main stream of the research effort of the Aviation Psychology Program, and, therefore, will be allowed to stand alone.

Discussion in this chapter will proceed under four major topics: organization and operating procedures, analyses of duties, criteria of proficiency and validation of stanines and test scores. While the obvious orientation of this section and of the work it reviews is toward evaluation and improvement of psychological selection procedures, the close connection between selection and training procedures can hardly be overemphasized. For example, while the criteria of proficiency were studied primarily to find scores against which to validate tests, they would often have been equally useful in evaluating training. For only insofar as measures of achievement existed, with which individuals were measured from time to time throughout a course of training, could the effectiveness of the efforts spent in training be critically appraised.

ORGANIZATION AND GENERAL OPERATING PROCEDURES

Certain comments need to be made on administrative experience gained in fulfilling the assigned missions in the four continental air forces. This experience revealed several factors which affected in an important way the accomplishment of research in the operational training situation.

Number of Research Personnel

In the first place, it seems almost self-evident that the number of psychologists assigned was insufficient for adequate completion of the prescribed mission. From the fall of 1944 until the close of the war, there were at most times roughly 4,000 combat crews in training at the approximately 80 CCTS and OTU stations in the 4 numbered air forces. There were in addition more than 20 stations devoted to the training of fighter pilots. It was impossible with the available personnel to make more than an occasional visit to most of these stations, so that detailed procedures in the assignment of grades and evaluation of performance could not often be followed up on the spot. There was thus no way research personnel could keep track of the frequent changes in training procedures or the special conditions which frequently arose in the training at a particular station.

Heterogeneity of training and the vast size of the air forces prevented aviation psychologists in three of them from maintaining complete coverage of all trainees being trained in their commands. These two factors were somewhat less in evidence in the First Air Force where no heavy bombardment training was done and where research personnel were able to con-

centrate upon a single type of training—P-47 fighter pilot training. It was thus possible in the First Air Force to keep a complete file of the P-47 trainees in the command. The problem could not be solved in the other three air forces by having a small staff at headquarters perform the evaluation studies, since proficiency records of individual trainees were not often transmitted from station to headquarters, and since the local stations had a great deal of autonomy in carrying out their training. Only at the station level were the actual names of crew members available prior to completion of operational training and assignment to staging for combat.

The small number of aviation psychologists available had another effect on research in the Continental Air Forces. Because there were so few of them, research personnel hesitated to undertake the tremendous job of instituting new procedures of evaluation at any large number of stations in the various air forces, and spent considerable time studying and attempting to use existing records and routine procedures, even when it was apparent that such records and procedures were far from ideal.

Allocation of Research Personnel

Since in the Second Air Force, the aviation psychologists were assigned to each training wing, while in the other air forces they were concentrated at a central point, some question arises as to which was the more effective organizational plan. In general, it is believed that concentration was more effective. However, there were some advantages to deployment in the wings. These included location in the geographical vicinity of a number of bases which carried on the same type of training, and direct contact with a headquarters primarily concerned with fulfilling training directives. On the other hand, day to day contact with each other in central units enabled research personnel to criticize and revise research procedures and to pool statistical services. Visits to bases were made on a temporary duty basis, and, while likely to be less frequent than those made from the wing, were probably equally effective. Continuous discussion at the central point made it possible to shift research emphases and to transfer work to those bases where conditions were most suitable to the research being undertaken.

Effecting Useful Research in Operational Training Situations

Throughout the psychological program, an attempt was continuously made to judge the importance of any research undertaking in terms of its possible contribution to the war effort. In order for the results of research to lead to action, an operating problem had to exist and be recognized by command and administrative personnel. The research then had to be directed fairly specifically toward a solution of that problem. In the case of studies on the validation of test scores and stanines the action was clear enough. Any new facts regarding the relative validities of tests could be readily used in revision of the classification test battery administered in the psychological units of the training command.

In other types of studies the mechanisms for getting action based on research results had to be worked out. For example, for a long time it was believed by various officers in the air forces that crews should be assembled before training on some sort of proficiency basis, and that information on the relative success of the whole crew in training should be transmitted to the overseas commander for use in determining suitable crew assignments. Some research had been done by psychologists on the characteristics of successful lead crews. However, research in methods of crew evaluation was not always supported locally and did not at first have any practical outcome. Later, psychologists had opportunity to make recommendations in a planning conference and aided in the preparation at Headquarters, Army Air Forces, of an AAF Letter outlining specific procedures to be used in assigning crews and directing that their proficiency be routinely evaluated at the station level. Since this AAF Letter required that an evaluation be sent overseas with each crew, it was possible to send through command channels a letter from Headquarters, Continental Air Forces, directing each member air force to undertake research on individual and crew proficiency measurement in order to establish suitable procedures for use in the evaluation called for in the AAF Letter. With this clear administrative need for a measuring instrument at the station level, psychologists were invited to work in close cooperation with training personnel on the solution of what was now a practical problem of immediate interest to all concerned. This experience illustrates how proper coordination of research objectives at all levels of commands often results in better recognition of the need for certain research and brings about closer cooperation among interested sections in the attainment of the research objectives.

ANALYSIS OF DUTIES

Job analysis is a technique or a tool which is almost always employed by the psychologist when faced with problems of selection or classification for jobs in industry. The type of analysis, and the level of detail in description of the duties depend upon what use is to be made of the information. Throughout the Aviation Psychology Program frequent use has been made of job analysis techniques. In the early days of the program, a series of bulletins on the analysis of duties of all aircrew positions was published for the guidance of research workers in devising tests and test procedures. Informal analyses were constantly being made as research workers came in contact with aircrew personnel on the job. The Psychological Research Projects for each aircrew position, located in the AAF Training Command, devoted considerable effort to systematic and relatively exhaustive job analyses of the aircrew position with which they were concerned, especially to the students' job in the early stages of training. Research workers in the Continental Air Forces, on the other hand, did not undertake extensive formal job analyses of aircrew duties since job requirements in operational training were essentially similar to those in the AAF Training Command,

although perhaps different in some details. However, reference was made throughout the text of this report to training manuals, etc., which outlined when and how various skills were to be exercised and what further training was given in each specialty in operational training.

In this connection, it might be argued that, in one sense, the studies of the criteria reported in the preceding chapters represent job analyses of a sort. The object of these efforts was to discover or to develop methods of measuring individual differences in skill or competence on the job, and depended for success upon how representative each variable was of the individual's skill in the total requirements of the job. Insofar as variables were discovered and measured in which individuals on the job did differ from one another in manner of performance, it may be said that an operational definition of the job was developed. Obviously, such criteria in the aggregate do not enumerate *all* the activities required nor *all* skills and aptitudes which must be present in at least minimal amounts for the individual to hold the job at all.

Some attempts at systematic job analyses were made for three types of aircrew duties in the training air forces. These analyses emphasized especially the phases of the job which were new to the trainee when he entered operational training from the AAF Training Command.

Fighter Pilot

The first of the more detailed analyses had to do with the job of the fighter pilot, that is, the job of the man who had just learned to fly an airplane and now had the job of learning to fight with it. The skills required were determined by interview, by examination of training directives, by study of ratings and objective criteria, and from analysis of records of accidents and Flying Evaluation Board proceedings. Traits, defined as those modes of behavior which appear to spring from the psychological background of the pilot, were determined chiefly by interviews with trainees, instructors and returned combat pilots. The general conclusion of these investigations was that within the Continental Air Forces almost all the students possessed the minimum flying skills, (perhaps because the weaker students were eliminated in the Training Command), so that many of the most important differences between good and poor pilots lay in the personality characteristics determining aggressiveness, ability to take responsibility and judgment.

Airplane Commander

The second area of job analysis was that of the airplane commander. Here attention was centered not only on the skills and traits required in flying a bombardment aircraft, but also in the requirements for the adequate exercise of command and leadership in a military situation. Two approaches were made: one through direct observation of the airplane

commander in the air on training missions, and the other through the collection of reports from returned combat personnel and trainees about situations in which airplane commander ability was exercised. The results of these studies point to the importance of certain personality characteristics which seem to determine success in leadership and command, characteristics which were assessed to only a very limited extent in preselection and classification procedures. The research workers also observed that relatively little training was given the airplane commander in the techniques of leadership or the ways of exercising command. It is the opinion of the writers that specific instruction can be given in these matters, and that it is essential to set up special instructional procedures in command techniques, utilizing the known principles of executive leadership combined with illustrations of the adequate exercise of command in aerial bombardment warfare. To young pilots, it is believed that much distilled experience in combat training leadership could be transmitted by precept and example through instruction specific to the duties and responsibilities of the airplane commander.

The Bombardment Crew

The third area for specific job-analysis was that of the crew-as-a-whole, with particular attention to the duties and responsibilities of the lead crew. Source material for these studies was confined to interviews with instructors and especially aircrew members returned from combat. Unanimous agreement was quickly reached on the importance of individual proficiency in determining the success of the crew. Also, the necessity for teamwork and cooperation among all members was heavily stressed. However entire agreement was not reached in the interesting question of whether a complete assessment of the individual proficiencies of each crew member would give an adequate measure of the proficiency of the crew as a whole. The question is an important one, and some opportunity for a quantitative answer to it was set up in the procedures established under AAF Letter 50-117. However, hostilities ceased and combat crew training was stopped before an answer could be obtained.

CRITERIA

Studies of the criteria of proficiency of individuals and crews in operational training constituted the most important activity of aviation psychologists in the Continental Air Forces. Reliable measurement of skill in job performance is, of course, fundamental to any adequate assessment of either selection or training procedures. In this section the criteria of proficiency studied are discussed briefly in terms of certain important characteristics and there is also brief discussion of some of the special conditions of operational training which influenced the outcome of studies made.

Classification of Proficiency Criteria

General Comments

Criteria of proficiency might be classified in an almost infinite number of different ways. The particular categories chosen for the analyses in this section are thus but a few out of many possible classifications. As will be obvious upon examination, the various categories of classification are not in practice by any means independent of one another. However, each dimension of classification given does represent a somewhat different emphasis, and an emphasis which is believed by the authors to be of interest in the consideration of the criteria.

Directness

The first dimension in terms of which the criteria are classified below is the directness-indirectness of the measurement of the skill involved. This dimension runs from those criteria which measure the actual performance of some aspect of the job of the aircrew member concerned, as for example, the number of hits that a fighter pilot gets in an aerial target, to some measure of an indirect outcome of skillful performance of the job such as rate of training accomplishment. Analysis in terms of this dimension is important in the selection of criteria for use in training, where the more direct a criterion is, the more effective it is likely to be in the learning situation.

Generality

A second continuum on which proficiency criteria may be placed might be called generality-specificity. It refers to the extent to which a measure gets at the total performance of an aircrew member on his job. For example, the records of landing and take-offs obtained from mobile control units represent specific or limited aspects of the job of fighter pilot, while ranking airplane commanders in order of expected proficiency in combat represents a very general measure.

While of two criteria otherwise equally desirable the more general would probably be best for use as a sole criterion, it would be quite possible for a combination of specific criteria to be superior to any available combination of general criteria. In this connection the question arises of how various specific measures are related to each other and how an estimate of an individual's over-all proficiency on the job can be most economically derived. A number of separate specific criteria are likely to be more useful in training than general criteria since they would reveal specific weaknesses in the results of training and would direct the attention of instructors and students thereto.

Simplicity

Precise measurement constitutes the cornerstone of any scientific edifice. In the last analysis, all measurement depends on a judgment made by an

observer. The object in developing precise instruments and techniques of measurement is to provide situations such that the judgments made by observers are as simple as possible. Reading the scale value to which an instrument needle is pointing, is one of the simpler types of judgment. Rating a personality characteristic of a pilot trainee is one of the more complex types. Thus one of the most important categories of classification of criteria in this report is that described in terms of the simplicity-complexity aspect of the judgmental situation.

Objectivity

There might be some question as to whether there be needed to consider the objectivity-subjectivity of criteria in addition to their classification in terms of simplicity. Certainly judgments made in simple situations are less likely to be influenced by subjective factors than those involving extreme complexity. However, it is believed that the emphasis is sufficiently different to warrant consideration of the various criteria in terms of objectivity as well as simplicity.

Permanence

The final dimension or category of classification of criteria, is that defined in terms of the permanence-transitoriness of the nature of the phenomena observed. Results of written proficiency examinations, the answer sheets of which might be held indefinitely, would represent relatively permanent criteria. The judgment by the mobile control unit operator regarding the quality of a particular landing of a plane would represent a criterion at the transitory end of the scale.

In table 14.1 the criteria used in the studies reported in this volume are roughly categorized in terms of the dimensions described above. Under each dimension, each criterion is given a value of "H," "M," or "L," representing high, moderate or little amount of the criterion quality represented by the first term of each pair describing the various dimensions. Thus in the column marked "Sim.-Comp." an "H" would mean a criterion such that the observation involved only a simple judgment while an "L" would refer to a criterion such that a complex judgment was involved. An "M" would refer to a criterion lying between the two extremes. It is recognized that the particular values listed for the various criteria in table 14.1 are somewhat arbitrary. However, they represent the composite judgment of several aviation psychologists who participated in the collection and analysis of much of the criterion data.

Also shown in the table are rough averages of the reliability coefficients for various proficiency criteria. The approximate numbers of cases included are shown in parentheses following each coefficient. Mission grades and comment scores are seen to have the highest degree of reliability of all the criteria studied. Gunnery scores, bombing scores and sortie grades have moderate reliability. With the exception of the ratings of individual

Table [illegible].—Classification of [illegible]

Criterion	Aircrew position	Dir.-ind.	Gen.-spe.	Sim.-comp.	Obj.-subj.	Per.-tran.	Inter-observer	Test retest	Odd-even missions
Gunnery and bombing:									
Air-to-air gunnery	FP	H	M	H	H	H			0.31 (1400)
Air-to-ground gunnery	FP	H	L	H	H	H			.30 (720)
Skip bombing	FP	H	L	H	H	H			.31 (130)
Dive bombing	FP	H	L	H	H	H			.40 (300)
Skeet scores	FP	H	L	H	H	H			.30 (100)
Average circular error	B, crew	H	H	H	H	H			.30 (600)
	APC	M	H	H	H	H			
	CP, N	L	H	H	H	H			
Radar bombing scores	N, B, crew	M	H	H	H	H			.45 (300)
	APC	M	H	H	H	H			
Gun camera scores	Gunners	M	H	H	H	H			.30 (350)
Mission performance:									
Mission grades	FP	M	M	L	M	L			.80 (300)
	N	M	M	L	M	M			.42 (150)
Mission comments	FP	M	M	M	M	L			.80 (150)
Mission ratings	APC, N, B, ROR, crew	M	L	M	M	L	0.45 (250)		.60 (150)
Mobile control reports	FP	M	M	M	M	L		0.30 (400)	
Crew proficiency check[1]	APC, crew	H	M	M	H	L			
Photo sortie grade	Pilot	H	M	M	M	M			.46 (150)
ETA error	N	M	M	M	M	H			.13 (200)
Course error	N	M	M	M	M	H			.65 (200)
Over-all rating:									
Numerical ratings	FP, APC, CP, B, N	L	H	L	L	L	.50 (300)	.45 (80)	
Rankings	FP, APC, CP, B, N	L	H	L	L	L	.60 (250)		
Instructor interviews	FP	L	H	L	L	L	.65 (650)		
Interview board rating	APC, CP, N, B	M	H	L	L	L			
Descriptive ratings	FP	L	H	L	L	L	.60 (725)	.35 (70)	.71 (720)
	APC, N, B, ROB, FL	L	H	L	L	L	.60 (1000)		
	Crew	M	H	L	L	L	.25 (250)		
	Gunners	L	H	L	L	L	.60 (160)		
Mutual ratings	FP, APC, N, B	M	H	L	L	L	.65 (130)		
Self ratings	APC, N, B	L	H	L	L	L			
Questionnaire[2]	APC	L	M	L	L	L			

[1] The within mission reliability obtained by correlation of odd vs. even item scores was .80.
[2] The internal consistency coefficient was .60, based on the Kuder-Richardson formula.

TABLE 14.1.—*Classifications and reliabilities of criteria (cont'd)*

Criterion	Aircrew position	Criterion characteristics: Dir.-ind.	Gen.-spec.	Sim.-comp.	Obj.-subj.	Per.-tran.	Reliability: Inter-observer	Test retest	Odd-even missions
Flying Evaluation Board action	FP, APC, N, B	L	M	L	L	H	....	....	....
Aircraft accidents	FP, APC	L	H	M	M	H	....	....	....
Training accomplishment:									
All requirements	FP, APC, N, crew, B	L	H	M	H	H	....	.00 (400)	.65 (1200)
Abortions	FP, APC, crew	L	M	H	H	H	....	....	....
Ground training:									
Academic grades	APC, CP, B, ROB, gunners	M	L	M	L	M	....	.50 (800)	....
	N, FE	M	M	M	M	M	....	....	....
	FP, APC, B	L	L	H	M	H	....	....	....
Examinations[1]	N	M	M	H	M	H	....	....	....
	APC	M	L	M	M	M	....	....	....
Synthetic trainers CNT—grades	N	M	M	M	M	M	....	....	.40 (1000)
CNT-ETA error	N	M	L	H	H	M	....	....	.40 (100)
CNT-Course error	N	H	L	H	H	M	....	....	.11 (100)
CNT-Number fixes	N	H	L	H	M	M	....	....	.50 (100)
DRT—grades	N, B	L	L	M	M	M	....	....	.30 (200)
Link—ratings	APC, CP	M	M	M	M	M	....	....	.30 (1000)
Link—Hours to finish	APC, CP	M	M	M	M	M	....	....	....
Bomb trainer 7-A-3	B	L	M	M	M	H	....	....	....
		M	H	M	M	M	....	....	.41 (250)
Type of assignment:									
As instructor	FP	L	H	L	L	L	....	....	....
As copilot or commander	APC, CP	L	H	L	L	L	....	....	....

[1] The within mission reliability obtained by correlation of odd vs. even item scores was .80.

here is m[illegible] between observers on the ratings of [illegible] various criteria of navigation proficiency, gun [illegible] scores and various ground trainer scores have a low degree of re[illegible].

Special Criterion Problems Encountered

Variability

Throughout the Continental Air Forces, there were great variations in training procedures, in methods of recording proficiency and in the interpretation of training requirements. There were differences at all levels of organization, between flights, classes, stations and air forces. Some of these differences, such as those having to do with scoring of bombing performance, naturally had a vital effect upon the scores of crews at certain stations. This variation in conditions not only served to attenuate any relationships between stanines and proficiency criteria, but also may have discouraged attempts of training personnel to develop better proficiency measures, since they knew that any such measures developed would probably be limited in use to the station where they were developed.

Lack of Interest in Proficiency Measurement

For a number of reasons there was little interest on the part of many training officers in the measurement of proficiency. Two of the most important reasons were as follows. First, there was no practical administrative procedure set up for eliminating individuals and crews thought not proficient. All administrative pressure was to graduate every crew regardless of proficiency, so as to meet personnel commitments to combat theaters. A second reason for the lack of interest shown was the belief that all of the individuals reaching operational training had attained a satisfactory degree of proficiency. Training officers frequently stated "all of our crews are proficient."

Problem of Control Groups

Because facilities were not available, the aviation psychologists in the various heavy bombardment commands were unable to maintain a complete file of names of all personnel entering operational training. It was thus necessary in some cases to choose a control group for comparison with an experimental group, or a group where a particular condition had obtained. Such controls almost inevitably turned out to have been open to some sort of bias, operating either explicitly or implicitly in their selection.

VALIDITY OF STANINES AND CLASSIFICATION TEST SCORES

Taken by position, there is no doubt that the stanines (B, N and P) and individual classification test scores better predicted the proficiency in operational training of fighter pilots than they did the proficiency of individ-

uals of any other flying specialty. Validity coefficients were larger and more consistently positive, both in different samples of a particular type of score, and among different types of scores. Three types of proficiency measures might be singled out as especially promising: gunnery scores (air-to-air and air-to-ground), mission comments, and training accomplishment scores. Stanine validities of roughly 0.15, 0.25 and 0.25 respectively, were obtained. While the gunnery and mission comment scores were also suitable for use in evaluation of training, the training accomplishment scores were not well suited for this purpose. For, if the rapid accomplishment of training requirements became a recognized goal and students were evaluated in those terms, other more important goals would be neglected.

Another important feature of the validity data presented in the foregoing chapters was the relatively poor success attained in the prediction of crew bombing accuracy. As with other criteria, the fact that the bombing scores were not perfectly reliable may have tended to obscure possible relationships. However, a second and more important consideration was the fact that bombing scores represented the cooperative performance of at least two and sometimes three aircrew members. Thus the stanines and classification test scores of a single aircrew officer could not be expected to account for more than a small part of the total variance in such scores. However, attempts to predict the crew bombing scores by a combination of stanines of several crew members did not yield promising results either. Here it is probable a number of additional sources of error were introduced since for only a few of the crews were stanines available for all aircrew officers. It is not known what selective factors were operating to determine which crew members were with and without such data.

Among the various types of criteria over-all ratings and descriptive ratings of all sorts were in general not well predicted by stanines of the individuals concerned. Somewhat more promising results were obtained for the ratings of performance on synthetic trainers. Link trainer scores of various types and CNT ratings were better predicted by the stanines than most other criteria. Finally, ground school grades usually reflected to some extent the stanines of the individuals concerned. This type of criterion suffers from the fact that it logically represents only a small part (if any) of the type of actvity actually carried on in combat operations. The "validity" of ground school performance is itself open to question, when the ultimate objective is the prediction of success in combat. Although few of the coefficients attained statistical significance, it can be pointed out in summary that positive results were obtained in most cases. Taking the data as a whole, individuals with higher stanines had some tendency to attain higher proficiency scores.

The validities of individual aircrew classification tests were obtained for only a few of the criteria studied and for the most part only with fighter pilots. No very striking results were obtained. In general, the Mechanical Principles and Mechanical Information Tests and the Biographical Data

best predicted success in operational training for fighter pilots. While most of these validity coefficients were positive, it is interesting that no one psychomotor test consistently yielded validity coefficients that reached statistical significance.

Two serious difficulties were encountered in validity studies in operational training. The first was the presence of differential factors of selection with respect to the original testing population. In a particular group of crews might be included individual crew members who belonged to testing population as much as two years apart. Not only had testing procedures and tests been revised, but the character of the population tested had changed. Because of the small numbers of cases available, it was frequently not possible to control this factor. The second difficulty was that introduced by extreme differences in level of experience of aircrew members in a single CCTS class. It is obvious that level of proficiency should vary with amount of experience. Thus, in validity studies utilizing populations of extreme variability in experience, it would be surprising if consistent correlations were found between original classification test scores or stanines and criteria of proficiency.

The data obtained do not permit a decision as to what might be the optimum combination of various more or less specific criteria. The degree of correlation between most of the proficiency measures studied was rather high considering the level of reliability of the various measures. Hence, whenever it was tried, the statistical combination of several measures yielded validities not much superior to those obtained with one or more of the separate scores which went into the combination. Most of the rating procedures would not lead one to expect that a "clinical" combination of criterion scores would give any higher coefficients of validity than did the direct statistical combinations. It is doubtful whether sufficient knowledge of the interrelationships of abilities exists to make possible a successful clinical combination of criteria. Ideally, for combination, one would probably desire a number of specific criteria, each of which measured a different aspect of the total job and the intercorrelations of which were low. Practically, this would ordinarily result in a combined criterion with a low degree of reliability. When there is added to this consideration the fact that the intercorrelations, and even frequently the reliabilities of the various possible scores, are not well established the decision as to what criteria or combinations of criteria should be used in validation becomes a matter for arbitrary decision using whatever insight the aviation psychologist may possess.

SUMMARY

Aviation psychologists assigned to the various Continental Air Forces spent approximately a year studying criteria of proficiency in operational training, determining validities of classification test scores and stanines in predicting success of aircrew officers in such training, developing procedures

for the identification of potential lead crews and performing various service activities of interest to the local commands. A relatively thorough survey was made of available records and procedures for evaluating proficiency of individuals and crews and additional procedures were developed in co-operation with training departments. Had training continued another 6 months, it would probably have been possible to put into universal operation some of the more promising of these procedures. As it was, only for fighter pilots was there available adequate information on proficiency for either validation of selection procedures or studies of effectiveness of alternative methods of training. Finally, within the limitations of the data there was a great deal of evidence that individuals with high stanines tended to reveal greater proficiency in operational training than did individuals with low stanines.

APPENDIX A

AAF Letter 20-103. Subject: Aviation Psychology in the Continental Air Forces

Appendix A.1

20-103

HEADQUARTERS, ARMY AIR FORCES
Washington, 19 July 1945.

AAF Letter 20-103,
SUBJECT: Aviation Psychology in the Continental Air Forces.
To: Commanding General, Continental Air Forces.

1. In accordance with AAF Regulation 20-59, aviation psychologists will be assigned to the Continental Air Forces, functioning under the technical supervision of the air surgeon, Continental Air Forces to carry out the following functions.

   a. Making available to commanding officers existing psychological test records, proficiency test scores, and training records, and advising concerning the use of such information in the selection of personnel for lead crew and other special types of operational training.

   b. Devising and/or administering aptitude and/or proficiency tests for the purpose of securing additional information to be used in the selection of personnel for special training.

   c. Developing measures and/or collecting data on the proficiency of individuals and/or combat crews to be used in evaluating the accuracy of selection procedures.

   d. Undertaking additional research studies that may be directed by the Commanding General, AAF or the Commanding General, Continental Air Forces.

2. Aviation psychologists will be attached to the surgeons of the several continental air forces and to the surgeons of such stations with these air forces as may be necessary to carry out the mission indicated above.

3. Aviation psychologists will have access to records of proficiency in operational training. They will be provided opportunities to gather supplementary information on aptitude and proficiency of personnel in operational training in so far as this does not interfere with the primary training mission of the Continental Air Forces.

4. Headquarters, Continental Air Forces will submit to the Commanding General, Army Air Forces, Washington 25, D.C., *Attention:* Air Surgeon, a monthly report of research activities of aviation psychologists in those air forces. The report for the month of June will include a summary report of research activities for the previous fiscal year. This report is assigned Reports Control Symbol AAF-AS-M-39.

BY COMMAND OF GENERAL ARNOLD:

[SEAL]

IRA C. EAKER,
Lieutenant General, United States Army
Deputy Commander, Army Air Forces

6-721, AF

APPENDIX B

Fighter Pilot Mission Records and Rating Scales

Appendix B.1

STUDENT MISSION RECORD

.... (Name)	.... (OSN)	.... (Date)
.... (Mission No.)		.... (Instructor)

Directions: Under each applicable heading write specifically what the student pilot did that was either (1) *Poor, faulty,* or *sloppy* (*particularly pointing out when the error made endangered his or other pilots' lives*) or (2) *Outstanding, remarkably good,* or *far above average,* particularly in view of his point in training. *Do not comment* when the pilot's performance was merely *satisfactory, adequate, fair, good but not much better,* or *passable.*

ADHERENCE TO BRIEFING INSTRUCTIONS:
BASIC FLYING TECHNIQUE (Taxi, take-off, traffic pattern, and landing):
JOIN-UP:
FORMATION:
NAVIGATION:
ACROBATICS:
COMBAT MANEUVERS:
INSTRUMENTS:
RADIO:
PERFORMANCE IN EMERGENCY SITUATION:
OTHER:

.... (Instructor's Initials)

Appendix B.2

STUDENT MISSION RECORD

Mission No. .... Date ....

Name .... OSN .... Instructor ....

Directions.—Opposite each applicable item check commendation or criticism in the appropriate column.
Particularly good means *outstanding,* in view of student's experience. *Weak* means *below standard,* in view of student's experience.
Do not check when the pilot's performance was merely *satisfactory, adequate, fair, good but not much better,* or *passable.*

	Weak	Particularly good
General:		
1. Adhering to briefing instructions		
2. Taxiing		
3. Take-off		
4. Traffic pattern		
5. Landing		
Formation flight:		
6. Joining formation		
7. Maintaining position		
8. Cross-overs		
9. Sighting other aircraft		
Navigation:		
10. Preparation for missions		
11. Pilotage (*Orientation*)		
12. Speed and accuracy of corrections (*in course, altitude and airspeed*)		
13. Adherence to ETA		
Other:		
14. Instrument flying		
15. Acrobatics		
16. Combat maneuvers		
17. Radio procedure		
18. Performance in emergency situation		
19. Air discipline and attitude		

Remarks.—

....
Instructor's initials

Appendix B.3

Proficiency rating scale for trainee fighter pilots at.................................

Full name........................... Rank.............. A.S.N..............

Class............................ Total number of flying hours.................

Arrival date......................... Departure date...........................

Directions. Instructors in consultation with their unit commanders will rate each of their trainees on proficiency as a fighter pilot. These ratings should represent the pooled judgment of all supervisors who have actually flown with the trainee. Trainees usually show greater proficiency in some aspects of flying than in others; so the ratings below should vary from one item to another. Trainee pilots are to be rated by *comparison with other men now in training, or having passed through in recent commitments.* Recent graduates from Training Command schools (newly-rated pilots) are to be compared only with other pilots with similar flying experience. In assigning ratings use should be made of *all five steps on the scale when they are applicable.* These ratings will not be seen by trainees, will not be entered on their records, and will not influence their future assignment in any way. Instructors will not, at any future date, be taken to task for any ratings they may have assigned.

Scale to be used for rating each item

(5) ...... Trainee shows *outstanding ability* on this item; is as good as best trainees now in training.

(4) ...... Trainee is *above average* on this item.

(3) ...... Trainee, on this item, falls in the majority or middle group of pilots; shows "average" ability.

(2) ...... Trainee is *slightly below average* on this item, compared with others now in training.

(1) ...... Trainee is weak on this item; proficiency not yet up to expected standards.

Rating of flying proficiency

[Encircle the appropriate number for each item below]

A. Basic flying technique:

1. Take-off	1 2 3 4 5
2. Flying in traffic pattern	1 2 3 4 5
3. Landing	1 2 3 4 5
4. Taxiing	1 2 3 4 5

B. Formation flying:

5. Joining formation	1 2 3 4 5
6. Maintaining position	1 2 3 4 5
7. Cross-overs	1 2 3 4 5
8. Use of throttle	1 2 3 4 5
9. Sighting other aircraft	1 2 3 4 5
10. Formation take-off	1 2 3 4 5

C. Navigation:

11. Preparation for missions	1 2 3 4 5
12. Pilotage (orientation)	1 2 3 4 5
13. Speed and accuracy of corrections (of course, altitude and air-speed)	1 2 3 4 5
14. Adherence to ETA	1 2 3 4 5

D. 15. Instrument flying (single over-all rating)	1 2 3 4 5
E. 16. Acrobatics (single over-all rating)	1 2 3 4 5
F. 17. Night Flying (single over-all rating)	1 2 3 4 5
G. 18. Combat flying (single over-all rating)	1 2 3 4 5
H. 19. Radio procedure (single over-all rating)	1 2 3 4 5

Rating of important pilot traits

1. Eagerness for combat duty (degree of enthusiasm for fighting)	1 2 3 4 5
2. Air discipline (dependability; how greatly would you desire to have this trainee flying on your wing in combat?)	1 2 3 4 5
3. Alertness (ability to learn quickly, remember previous instructions, and keep "on the ball" while flying)	1 2 3 4 5
4. Aggressiveness (zest shown in acrobatics, combat flying, and "on the deck" flying)	1 2 3 4 5
5. Headwork (ability to plan ahead and to foresee consequences; does trainee *think* while in the air?	1 2 3 4 5
6. Coolness (tendency to remain calm and unexcited when the unusual occurs)	1 2 3 4 5
7. Conscientiousness (desire to fly additional missions in the interest of self-improvement; does trainee study material he *might* need in addition to that he *must know?*)	1 2 3 4 5

Date of this proficiency rating ....

Signature of instructor ....

Initials of unit or flight commander ....

Remarks ....

....

....

....

....

Appendix B.4.

Proficiency rating scale for training fighter pilots at ....
Full name .... Rank............... A.S.N.............
Class............................ Total number of flying hours.................
Arrival date........................ Departure date........................

Directions. Instructors in consultation with their unit commanders will rate each of their trainees on proficiency as a fighter pilot. These ratings should represent the pooled judgment of all supervisors who have actually flown with the trainee. Trainees usually show greater proficiency in some aspects of flying than in others; so the ratings below should vary from one item to another. Trainee pilots are to be rated by *comparison with other men now in training, or having passed through in recent commitments.* Recent graduates from Training Command schools (newly rated pilots) are to be compared only with other pilots with similar flying experience. In assigning ratings use should be made of *all five steps on the scale when they are applicable.* These ratings will not be seen by trainees, will not be entered on their records, and will not influence their future assignment in any way. Instructors will not, at any future date, be taken to task for any ratings they may have assigned.

Scale to be used for rating each item

(5) ...... Trainee shows *outstanding ability* on this item; is as good as best trainees now in training.
(4) ...... Trainee is *above average* on this item.
(3) ...... Trainee, on this item, falls in the majority or middle group of pilots; shows "average" ability.
(2) ...... Trainee is *slightly below average* on this item, compared with others now in training.
(1) ...... Trainee is weak on this item; proficiency not yet up to expected standards.

Rating of flying proficiency

[Encircle the appropriate number for each item below.]

A. Formation flying
1. Maintaining position 1 2 3 4 5
2. Use of throttle 1 2 3 4 5
3. Sighting other planes 1 2 3 4 5
4. Formation at high altitude 1 2 3 4 5

B. Joining Formation:
5. Joining as quickly as possible 1 2 3 4 5
6. Getting into and maintaining a correct position 1 2 3 4 5

C. Maneuvers:
7. Doing maneuvers difficult to follow 1 2 3 4 5
8. Doing tight turns 1 2 3 4 5

D. Gunnery Pattern:
9. Position before starting pass 1 2 3 4 5
10. Curve of pursuit 1 2 3 4 5
11. Rejoin 1 2 3 4 5

E. Aerial Gunnery:
12. Firing in range 1 2 3 4 5
13. Line of sight 1 2 3 4 5

F. Ground Gunnery:
14. Pattern 1 2 3 4 5
15. Firing in range 1 2 3 4 5

Rating of important pilot traits

1. Eagerness for combat duty (degree of enthusiasm for fighting) ... 1 2 3 4 5
2. Air discipline (dependability; how greatly would you desire to have this trainee flying on your wing in combat?) 1 2 3 4 5
3. Alertness (ability to learn quickly, remember previous instructions, and keep "on the ball" while flying) 1 2 3 4 5
4. Aggressiveness (zest shown in acrobatics, combat flying, and "on the deck" flying) 1 2 3 4 5
5. Headwork (ability to plan ahead and foresee consequences; does trainee *think* while in the air?) 1 2 3 4 5
6. Coolness (tendency to remain calm and unexcited when the unusual occurs) 1 2 3 4 5
7. Conscientiousness (desire to fly additional missions in the interest of self-improvement; does trainee study material he *might* need in addition to that he *must* know?) 1 2 3 4 5

Date of this proficiency rating ....

Signature of instructor ....

Initials of unit or flight commander ....

Remarks ....

....

....

....

....

APPENDIX C

B-29 Airplane Commander Questionnaire

Appendix C.1

B-29 Airplane Commander Questionnaire

Instructions

This questionnaire is being filled out for the Research Section of the Third Air Force. Its purpose is to gain a better idea as to what the B-29 Airplane Commander does and does not do in his relation with his crew. It will also serve as a guide for the Air Force in preparing future courses, checks and aids for the B-29 Airplane Commander.

The questionnaire is in no way to be considered a test or an examination. There are no right or wrong answers to the questions. The only right answer is the one that applies to you or best describes your actions.

The answers you give will in no way affect your present status nor will it have any bearing upon your present or future standing as an airplane commander. The results of the questionnaire will not be given to the training or directorial personnel of this base but will be used only by the Third Air Force Research Section. Its purpose is to gain information about and a better understanding of the airplane commander's problems.

Remember, there are no right or wrong answers. The only right answer is the one that best describes your situation or actions.

....
(Last name) *(First name)* *(MI)* *(ASN)* *(Rank)*

....
(Field or Base) *(Class number)* *(Number weeks in training)*

....
(Total flying hours) *(Total hours B-29)*

Former flying instructor: Yes .... No. .... If "yes", in what type aircraft .... length of time as instructor (years and months) ....

Combat experience: Yes .... No. .... If "yes," fill in the following:

Theater ....

Type aircraft flown in combat ....

Number Missions .... (or hours ....).

Airplane Commander Inventory

1. Have you witnessed or heard any dissension among your crew members during the past week?
   ...... a. Have witnessed or heard considerable dissension in the past week.
   ...... b. Have witnessed or heard some crew dissension in the past week.
   ...... c. Have witnessed or heard no crew dissension in the past week.
2. How much off-duty time have you spent with your officer crew members during the past week?
   ...... a. Have spent as much time as possible on off-duty hours with officer crew members during the past week.
   ...... b. Have spent considerable off-duty time with officer crew members during the past week.

...... c. Have spent some off-duty time with officer crew members during the past week.

...... d. Have spent very little off-duty time with officer crew members during the past week.

...... e. Have spent no off-duty time with officer crew members during the past week.

3. Do you feel it is important and desirable to associate socially with your officer crew members?

...... a. Feel it is extremely important and do so at every opportunity.

...... b. Feel it is important and associate with them a great deal.

...... c. Feel it is of little importance and association should be minimized.

...... d. Feel it is somewhat detrimental and such association should be avoided.

4. Check opposite the positions listed below the crew members whose home town you know.

...... a. Copilot.

...... b. Navigator.

...... c. Bombardier.

...... d. Radar operator.

...... e. Flight engineer.

...... f. Radio operator.

...... g. Right scanner.

...... h. Left scanner.

...... i. Tail gunner.

...... j. Ring gunner.

5. Have you fulfilled all promises you have made to your crew during training at this base?

...... a. Have fulfilled all promises I've made to my crew.

...... b. Have usually fulfilled promises made to my crew.

...... c. Have occasionally fulfilled promises I've made to my crew.

...... d. Have made no promises to my crew.

6. Do you feel it is desirable to associate socially with the enlisted members of your crew?

...... a. Feel it is desirable to associate socially with enlisted members of my crew and do so frequently.

...... b. Feel it is desirable to associate socially with enlisted members of my crew and do so occasionally.

...... c. Feel it is desirable to associate socially with enlisted members of my crew but do not do so.

...... d. Feel it is not desirable to associate socially with enlisted crew members but do so frequently.

...... e. Feel it is not desirable to associate socially with enlisted crew members but do so occasionally.

...... f. Feel it is not desirable to associate socially with enlisted crew members and refrain from doing so.

7. How have you dealt with problems involving crew difficulties or incompatibilities during your training at this base?

...... a. Have turned over to higher authorities all difficulties which have arisen.

...... b. Have turned over to higher authorities only those problems and difficulties which I have been unable to correct myself.

...... c. Have been unable to solve some difficulties myself but have not turned them over to higher authorities.

...... d. Have suspected problems and difficulties existed but did not investigate further thinking they would solve themselves.

...... e. I am certain no problems or difficulties have occurred with my crew.

8. Is there a "trouble maker" on your crew?

...... a. Yes.

...... b. No. If "yes," describe in what way he causes trouble ....

....

....

9. **Do you feel it is desirable to have dinners or social engagements at which entire crew (both officers and enlisted men) is present?**
   ...... a. **Feel social engagements for entire crew are desirable and should be held frequently.**
   ...... b. **Feel social engagements for entire crew are desirable and should be held occasionally.**
   ...... c. **Feel social engagements for entire crew are somewhat desirable but should be held very infrequently.**
   ...... d. **Feel social engagements for entire crew are detrimental and such engagements should not be held.**

10. **Have any of your crew members been involved in disciplinary action (i.e., official reprimand by an officer who is not a member of your crew; A.W. 104; courts martial) during your training at this base?**
   ...... a. **Three or more crew members have been involved in disciplinary action during our training at this base.**
   ...... b. **Two crew members have been involved in disciplinary action during our training at this base.**
   ...... c. **One crew member has been involved in disciplinary action during our training at this base.**
   ...... d. **No crew member has been involved in disciplinary action during our training at this base.**

11. **After each of the past four missions have you discussed, reviewed or critiqued with your crew, errors or problems which occurred during the mission?**
   ...... a. Yes; held a critique after each of the past four missions.
   ...... b. Yes; held a critique after three of the past four missions.
   ...... c. Yes; held a critique after two of the past four missions.
   ...... d. Yes; held a critique after one of the past four missions.
   ...... e. Have held no critique after the last four missions.

12. In your critique of a mission do you usually:
   ...... a. Talk to the entire crew as a group.
   ...... b. Talk with some but not all crew members as a group.
   ...... c. Talk with individual crew members only.
   ...... d. Do not hold a critique.

13. In the last 4 personal equipment inspections of your crew did you usually:
   ...... a. Make a thorough, detailed inspection of all crew members.
   ...... b. Make a thorough inspection for only certain crew members.
   ...... c. Make a rapid and superficial inspection of all crew members.
   ...... d. Fail to make a personal equipment inspection.

14. Have you designated to individual crew members responsibilities and duties in addition to those ordinarily required for the performance of their assigned positions?
   ...... a. Yes; to a very few crew members (1 or 2).
   ...... b. Yes; to some crew members (3, 4 or 5).
   ...... c. Yes, to half or more (6, 7 or 8).
   ...... d. Yes, to most crew members (9 or 10).
   ...... e. No; have not designated additional duties and responsibilities to any crew members.

15. To whom do your crew members bring their personal problems?
   ...... a. To you as the airplane commander.
   ...... b. To the copilot.
   ...... c. To the bombardier or navigator.
   ...... d. Other (name by position ....).
   ...... e. Personal problems are brought to no crew member's attention in particular.

16. When necessary to correct a crew member regarding the performance of his duties do you generally:

...... a. Take up the matter immediately regardless of place or time of occurrence.

...... b. Take up the matter at a time when the majority of the crew are present.

...... c. Have a private talk with the crew member involved.

...... d. Have had no occasion to correct a crew member regarding the performance of his duties.

17. Do you keep personal records of individual proficiency of all crew members?

...... a. Keep adequate records of all crew members.

...... b. Keep incomplete records which are usually sufficient for my purpose.

...... c. Keep records but they are not very good.

...... d. Would like to keep records but have been too busy to do so.

...... e. Do not keep records.

...... f. Do not keep records and do not think records are necessary.

18. Below is a list of sources from which you can learn about the proficiency of your crew members. Check as many of these sources as you have used during your training at this station:

...... a. Ground school records (engineering, grades, class records)

...... b. Trainer records (A-5, supersonic, etc.)

...... c. Flight records (logs, mission reports, etc.)

...... d. Instructor opinion.

...... e. Talking with crew members.

...... f. Observing crew members at work.

...... g. Other, describe.

...... h. Have checked none of the above.

19. When you receive extra or incidental information (from other than classroom sources) which is related to or would assist in the training of your crew, do you

...... a. Pass it on to the entire crew during a "bull session" or crew meeting.

...... b. Pass it on to the crew member to whose job the information is most closely related.

...... c. Keep the information for your own personal use.

20. Prior to a mission, do you give to your crew instruction and/or orientation in addition to that given at briefing?

...... a. Always give additional instruction and/or orientation to crew prior to a mission.

...... b. Nearly always give additional instruction and/or orientation to crew prior to a mission.

...... c. Frequently give additional instruction and/or orientation to crew prior to a mission.

...... d. Occasionally give additional instruction and/or orientation to crew prior to a mission.

...... e. Seldom give additional instruction and/or orientation to crew prior to a mission.

...... f. Do not give additional instruction and/or orientation to crew prior to a mission.

21. In regards to giving your crew briefing and/or instruction additional to that given in regular briefing, do you feel

...... a. You should always give additional briefing and/or instruction.

...... b. You should frequently give additional briefing and/or instruction.

...... c. You should occasionally give additional briefing and/or instruction.

...... d. You should never give additional briefing and/or instruction.

22. Select the statement below which best describes the readiness and storage of equipment prior to take-off, on your last four missions.
...... a. All equipment was ready and properly stowed.
...... b. All equipment was present but not properly stowed.
...... c. Not all equipment was ready but that which was ready was properly stowed.
...... d. No check was made.

23. When you have extra "waiting time" between briefing and take-off, do you
...... a. Utilize this "waiting time" by instructions, practice of training procedure, and/or discussions of some phase of training.
...... b. Utilize this "waiting time" by holding informal "bull sessions" in which entire crew participates.
...... c. Keep all the crew in the vicinity of the plane unless I specifically authorize them to leave but do not utilize this "waiting time."
...... d. Allow each crew member to choose his own activity until it is time to take-off.

24. When crew members fail to use correct interphone procedure do you
...... a. Correct the error immediately when it occurs.
...... b. Wait until you are on the ground to correct the error.
...... c. Never bother to correct the error, because, on the whole, the crew's procedure is superior.
...... d. Allow the error to pass uncorrected because in most cases it is a slip that will be corrected with additional practice.
...... e. Never pay attention to the error so long as the message is heard clearly and distinctly.

25. During flight do you
...... a. Make frequent combat crew position reports.
...... b. Occasionally call for combat crew position reports.
...... c. Seldom call for combat crew position reports.
...... d. Do not call for combat crew position reports.

26. On the last four missions have you practiced emergency procedures where possible whether specifically called for by mission requirements?
...... a. Yes; on all four missions.
...... b. Yes; on three of the last four missions.
...... c. Yes; on two of the last four missions.
...... d. Yes; on one of the last four missions.
...... e. No; not on any of the last four missions.

27. Do you check to see that all equipment is kept in an orderly manner during flight?
...... a. I make frequent checks to see that all equipment is kept in an orderly manner during flight.
...... b. I occasionally make checks to see that all equipment is kept in an orderly manner during flight.
...... c. I seldom make checks to see that all equipment is kept in an orderly manner during flight.
...... d. I do not check to see that all equipment is kept in an orderly manner during flight.

28. Do you require crew members to report and/or record important land observations made during flight (e.g., villages, airfield with type of plane on field, factories, or industrial areas).
...... a. Require crew members to report and/or record all land observations during flight.
...... b. Specify to crew what type of land observations which are to be reported and do not require them to report anything else.
...... c. Do not require crew to make or report land observations.

29. Which of the following statements characterizes your procedure following the completion of the last 4 or 5 missions?

...... a. Held a critique with every crew member present.
...... b. Held a critique with the majority of crew members present.
...... c. Held a critique individually with some of the crew members.
...... d. Held a critique only with individuals who made errors.
...... e. Usually felt no critique was necessary.

30. Which of the following statements best describe your attitude and practice relative to maintaining full adherence to rules and regulations governing the conduct and actions of military personnel.

...... a. Require strict adherence at all times in the air and on the ground.
...... b. Require strict adherence at all times in the air and moderate on the ground.
...... c. Strict adherence at all times in the air but feel adherence on the ground not essential.
...... d. Feel moderate adherence both in the air and on the ground is all that is necessary.
...... e. Feel that neither moderate or strict adherence is necessary in training while in this country.

31. Upon the completion of a successful mission do you

...... a. Give praise to the crew as a whole.
...... b. Give praise to those individuals whose work was outstanding.
...... c. Infrequently give praise because I feel it does no good.
...... d. Never give praise because I feel it is detrimental.

32. Which of the following statements best describe the way the enlisted crew members usually address you?

...... a. First name or nickname (such as Bill, Joe, Red, etc.)
...... b. Enlisted crew members address me by my last name only (such as Brown, Smith, etc.)
...... c. Enlisted crew members address me by my rank only (such as lieutenant, captain, major).
...... d. Enlisted crew members address me as "sir," or by using my rank and name (such as Lieutenant Brown, Captain Smith, Major Brown, etc.)

33. Which of the following statements best describe the way the officer crew members usually address you?

...... a. Officer crew members address me by my first name or a nickname (such as: Joe, Bill, Red, etc.).
...... b. Officer crew members address me by my last name only (such as Brown, Smith, Jones).
...... c. Officer crew members address me by rank only (such as lieutenant, captain, major, etc.).
...... d. Officer crew members address me as "sir," or by using my rank and name (such as Lieutenant Brown, Captain Smith, etc.).

34. Have you spent extra time in the past week with any crew member in helping him increase his proficiency?

...... a. Have had no occasion to do so.
...... b. Have not had any time to assist crew members.
...... c. Have helped one or two crew members.
...... d. Have helped three or more crew members.

35. Do you feel that an airplane commander should admit his mistakes to his crew?

...... a. Yes; his mistakes should be handled in the same manner as the mistakes of any of his crew members.

...... b. Yes; but his mistakes should be admitted to the officer crew members only.

...... c. Yes; he should admit his mistakes but a proper excuse should be made so that the prestige of the plane commander will not be affected by the incident.

...... d. No; the plane commander should not admit his mistakes because to do so would be detrimental to his control over the crew.

36. Check the following B-29 air crew positions you have flown in during your training at this base for purpose of orientation or familiarization.

...... a. Bombardier.
...... b. Navigator.
...... c. Flight engineer.
...... d. Radar operator.
...... e. Radio operator.
...... f. Scanner.
...... g. Tail gunner.

37. During critique and in discussions prior to a mission do you ask for questions, suggestions, and/or opinions of other crew members relative to the completion of the mission?

...... a. Have not done so because all information pertinent to the mission is covered by briefing.

...... b. Have occasionally asked crew members for questions, suggestions and/or opinions.

...... c. Have frequently asked crew members for questions, suggestions and/or opinions.

...... d. Have nearly always asked crew members for questions, suggestions and/or opinions.

...... e. Have always asked crew members for questions, suggestions, and/or opinions.

38. Have you had your crew members change positions during flight in order to get experience in other flight positions?

...... a. Have had all crew members change positions at least once and most have changed positions several times.

...... b. Frequently have crew members change positions when it does not interfere with the execution of the mission.

...... c. Occasionally have crew members change positions when it does not interfere with the execution of the mission.

...... d. Seldom have crew members change positions when it does not interfere with the execution of the mission.

...... e. Have not had crew members change positions but they have done so of their own accord.

...... f. Have not had any interchanging of positions on my crew as I feel it has not been necessary.

39. Would you request the transfer from your crew any man who (check any that apply).

...... a. Questioned your authority.
...... b. Was a source of crew dissension.
...... c. Was absent too frequently (even though the reason for his absence is justifiable).
...... d. Did not do his share of the crew's work in preparing for a flight and returning equipment after the flight.
...... e. Was not proficient in his specialty.
...... f. Was objectionable because of his personality.

40. What is the most difficult aspect of your job as an Airplane Commander? Write your answer below:

APPENDIX D

Pilot's Rating Sheet

Appendix D.1

PILOT'S RATING SHEET

MAAF Form 50-21A
(Rev. 5 June 45)

Pilot's name .... Crew No. ...... Mission No. ...... Date ....

Under "Rating" place the appropriate grade (5, 4, 3, 2, or 1) which you feel adequately describes the pilot according to the following rating scales: Exceptional 5; Above average, 4; Average, 3; Below average, 2; Poor, 1.

Then circle the letter (a, b, g, etc.) preceding those items which have influenced you in your determination of a grade, bearing in mind that you should compare him with your conception of the ideal lead crew pilot.

Check list	Rating	Remarks
1. Preflight:		
a. Fuselage		
b. Flight controls		
c. Fuel tank caps		
d. Control surfaces		
e. Landing gear		
f. Wheels and brakes		
g. Nacelles and engines		
h. Turbos and cowling		
i. Forms and reports		
2. Crew inspection:		
a. Parachute		
b. Clothing		
c. Mae West		
d. Crew briefing		
3. Air discipline:		
a. Interphone		
b. Crew coordination		
c. Instructions		
d. C. F. C. procedure		
e. Check lists		
f. Position identification		
4. Mission performance:		
a. Adherence to briefed plan		
b. Pilot-crew coordination		
c. Ability of pilot		
d. C-1 operation		
e. Target area		
f. Course and turns		
g. Airspeed and altitude		
h. Mission in general		
i. Return to base		
5. Airplane condition:		
a. Ash trays, traps etc.		
b. Locks—pilot covers		
c. Interior cleanliness		
6. Personal remarks:		
a. Initiative		
b. Crew respect		
c. Attitude		
d. Personality		

Appendix D.2

Over-all Rating of Pilot and Crew

In your over-all rating of pilot and of the crew it is essential that you think of them in terms of their ability as lead pilots. Circle the number above the appropriate description of the pilot or crew.

Instructor rating of pilot

5	4	3	2	1
Exceptional pilot. Outstanding performance. Recommended without qualification for lead crew.	Above average. Qualified for a lead crew. However additional effort would improve proficiency.	Average pilot. Needs brushing up. Shows possibilities for lead crew with proper direction and guidance.	Below average. Considerable progress and improvement necessary to qualify him for lead crew.	Poor pilot. Practically hopeless for lead crew without tremendous improvement.

Remarks.—

Instructor rating of crew

5	4	3	2	1
Exceptional discipline, coordination, interphone procedure and execution of briefed plan. Outstanding performance. Recommended without qualification as a lead crew.	Above average. Qualified as a lead crew. However additional effort would improve proficiency.	Average discipline, coordination, interphone procedure and execution of briefed plan. Needs brushing up. Shows possibilities as lead crew with proper direction and guidance.	Below average. Considerable progress and improvement necessary to qualify as lead crew.	Poor discipline, coordination, interphone procedure and execution of briefed plan. Practically hopeless as lead crew without tremendous improvement.

Remarks:

....

Instructor signature

APPENDIX E

Scale of Instrument Flying Skill—HB(B-17)

Appendix E.1

SCALE OF INSTRUMENT FLYING SKILL—HB(B-17)

(Experimental Form)

Directions for using the scale. In order to achieve standardized measurement, it is essential that directions be followed closely. Elements for each maneuver are listed. Check pilots will make sure pilot understands all the elements of a maneuver before beginning it, otherwise variations in testing will occur.

Turbulence may have a marked effect upon a pilot's execution of a maneuver. Check pilot will record his estimate of the amount of turbulence, but will base judgment of proficiency on behavior of pilot at times when turbulence is not effective. Insofar as possible, the proficiency grade should not be influenced by turbulence. By light turbulence is meant that the air is slightly rough but not rough enough to influence the pilot's performance. Moderate turbulence is when the air is rough enough that the pilot's performance of the maneuver is affected but the effect of the turbulence can be subtracted in evaluating the performance. Heavy turbulence is when the air is so rough that its effect on the pilot's performance cannot be estimated and the entire check on the given maneuver should not be used.

Proficiency in control of deviation is to be measured by indicating that type of correction used by the pilot during a specific time interval during a maneuver. Different characteristics of the flying will be observed during different time intervals. One of the following types of correction is to be recorded for each characteristic:

a. *Failure* to attempt correction of a deviation.

b. *Oscillating over-correction* or over-correcting in both directions resulting in an oscillating deviation.

c. *Simple over-correction* in recovery which sets up a deviation in the opposite direction; and *Simple under-correction* in which the original deviation is never entirely removed.

d. *Perfect flying*—no deviation occurring during interval.

Example. Spiral Climb. Performance of the pilot during climb will be measured during first 45 seconds after instructions on maneuver are given, using following sequence:

Instruction completed	First 15 seconds	Second 15 seconds	Third 15 seconds	Climb Completed
	No measurement	Bank deviation	Climb altitude	

During second 15-second period, *Bank deviation only* is to be evaluated. That type of correction used by the pilot most of the time during this 15-second period is to be recorded. For example, if the pilot uses oscillating over-corrections most of the 15-second period, oscillating over-correction will be checked in recording the pilot's performance.

During the third 15-second period, *Climb altitude deviation only* is to be evaluated.

That type of correction used by the pilot during his 15-second period in maintaining the correct attitude will be used evaluating his performance.

Proficiency evaluations as described above will be made of deviations in heading, airspeed, altitude, rate of turn, rate of climb, and rate of glide. Deviations in climb attitude, glide attitude, and bank will be measured on the flight indicator *only*.

Maximum deviations will also be recorded for some characteristics. This score is to be the largest deviation occurring in that interval during which the given characteristic was being observed. Deviations occurring at other times than during the specified interval are not to be recorded.

Degrees of heading error. During turns and glides the initial and final headings will be recorded in order to obtain the degree of error in perfomance.

Check pilot will score proficiency of pilot at completion of each maneuver. When possible at this time, pilot will be instructed to fly straight and level until scoring is completed and next maneuver is described. To score proficiency, place a check mark in the appropriate space for each measure of the maneuver. The check pilot will fill in the following information before the flight begins:

Pilot's name .... Rank .... ASN ....

Number of graduating class in Training Command .... Hours in B–17 ....

Pilot's status: Trainee .... Flying Inst. .... Administrator of Ground Inst. ....

Check pilot's name .... Rank ....

Date .... Time of day check began ....

On completion of flight check pilot will fill in the following:

Time of day check ended;

Over-all rating of pilot's instrument flying skill:

Excellent .... Good .... Poor ....

(*10% of pilots*) (*Majority of pilots*) (*10% of pilots*)

1. *Instrument take-off:*

Directions to check pilot. Align the airplane with the runway. With the pilot listening in, contact the tower and obtain and record the direction of any cross wind and the speed of the wind.

Cross wind at take-off: Direction ....° Speed .... m.p.h.

Runway reading ....°.

Instructions to pilot: Set directional gyro to nearest 5° indice of runway heading and take-off.

Rate of gaining directional control:

Gained directional control quickly ....

Gained directional control in average time ....

Gained directional control slowly ....

Altitude at precise moment that speed of 140 MPH is first attained .... feet.

Assistance on take-off run:

No assistance given ....

Assistance by words or signs ....

Assistance by movements of any controls ....

2. *Spiral climb* to right for 1,000 feet, fly straight for 30 seconds and climb to left for 1,000 feet.

a. Bank 15°.

b. Airspeed 150 m.p.h.

c. Rate 500 ft. per. min.

Check Pilot will have pilot put plane on a round numbered altitude e.g., 1,000, 2,000, 3,000, etc., and will inform pilot of the altitude 1,000 feet higher at which the climb is to be completed. At end of right climb and before beginning left climb pilot is to fly straight for about 30 seconds Altitude at end of each climb is taken at that moment wings are made level regardless of climbing attitude of plane

Bank deviation **and** ***Climb-attitude deviation,*** **are each to be observed during specified *fifteen* (15) second intervals in the following sequence:**

Instructions Completed	First 15 seconds	Second 15 seconds	Third 15 seconds	Turn Completed
	No measurement	Bank deviation	Climb attitude	

Maximum deviations **are to be recorded for each characteristic for the specified interval *only*, using the flight indicator. Deviations in bank are to be measured in degrees from the original bank set up by the pilot. Deviations in climb-attitude are to be measured in terms of the linear distance that the horizon line fluctuates from the original position set up by the pilot.**

Right turn

Altitude **at beginning of right turn .... ft.**

Altitude **at moment wings are first made level .... ft.**

Bank deviation **(2d 15-second interval only—check one):**

- **Failed to correct ....**
- **Oscillating over-correction ....**
- **Simple over or under correction ....**
- **Perfect flying—no deviation occurred ....**

Maximum deviation**—bank (Check one):**

- **± 2° limits ....**
- **± 4° limits ....**
- **± 6° limits ....**
- **Over ± 6° limits ....**

Climb attitude deviation **(3d 15-second interval only—check one):**

- **Failed to correct ....**
- **Oscillating over-correction ....**
- **Simple over or under correction ....**
- **Perfect flying—no deviation occurred ....**

Maximum deviation—climb attitude **(check one):**

- **± 1/16-inch limits ....**
- **± ⅛ inch limits ....**
- **± ¼-inch limits ....**
- **± ½-inch limits ....**

Altitude at beginning of left turn .... ft.

Altitude at moment wings are first made level .... ft.

Bank deviation (2d 15-second interval only—check one):

- Failed to correct ....
- Oscillating over-correction ....
- Simple over or under correction ....
- Perfect flying—no deviation occurred ....

Maximum deviation—bank (check one):

- ± 2° limits ....
- ± 4° limits ....
- ± 6° limits ....
- Over ± 6° limits ....

Climb attitude deviation (3d 15-second interval only, check one):

- Failed to correct ....
- Oscillating over-correction ....
- Simple over or under correction ....
- Perfect flying—no deviation occurred ....

Maximum deviation—climb attitude (check one):

- ± 1/16-inch limits ....
- ± ⅛ inch limits ....
- ± ¼-inch limits ....
- ± ½-inch limits ....

Trimmed plane (for whole maneuver): Yes ....; No ....

Weather during spiral climb:

Smooth ....

Light turbulence ....

Moderate turbulence ....

Heavy turbulence ....

3. *Level flight for 5 minutes*

a. Airspeed 155 m. p. h.

b. Heading ....° (to be given).

c. Altitude .... ft. (to be given).

Altitude deviation, heading deviation and air speed deviation are each to be observed during specified half minute intervals in the following sequence:

Instructions Completed	First half minute	Second half minute	Third half minute	Fourth half minute	Maneuver Completed
	No measurement	Altitude	Heading	Air speed	

Maximum deviations are to be recorded for each characteristic during the specified interval *only*.

Altitude deviation (2d half minute only—check one):

Failed to correct ....:

Oscillating over-correction ....

Simple over or under correction ....

Perfect flying—no deviation occurred ....

Maximum deviation—altitude .... ft.

Heading deviation (3d half minute only—check one):

Failed to correct ....:

Oscillating over-correction ....

Simple over or under correction ....

Perfect flying—no deviation occurred ....

Maximum deviation—heading .... degrees.

Air speed deviation (4th half minute only—check one):

Failed to correct ....:

Oscillating over-correction ....

Simple over or under correction ....

Perfect flying—no deviation occurred ....

Maximum deviation—air speed .... m. p. h.

Trimmed plane: Yes ....; No ....

Weather during flight:

Smooth ....

Light turbulence ....

Moderate turbulence ....

Heavy turbulence ....

4. 90° and 120° turns—90° turn to the right; straight flight for 30 seconds then 120° turn to left.

90° turn

a. 90° turn—bank 15°.

b. Airspeed 155 m. p. h.

c. Altitude .... ft. (to be given).

Bank deviation and altitude deviation are each to be observed during specified *ten* (10) second intervals in the following sequence:

Instructions Completed	First 10 seconds	Second 10 seconds	Third 10 seconds	Completed
	No measurement	Bank deviation	Altitude deviation	

Maximum deviations are to be recorded for each characteristic for the specified interval *only*.

Heading at beginning of 90° turn .... degrees.

Heading at moment wings are first made level .... degrees.

Bank deviation (2d 10-second interval—check one):

Failed to correct ....
Oscillating over-correction ....
Simple over or under correction ....
Perfect flying—no deviation occurred ....

Maximum deviation—bank (check one):

± 2° limits ....
± 4° limits ....
± 6° limits ....
Over ± 6° limits ....

Altitude deviation (3d 10-second interval—check one):

Failed to correct ....
Oscillating over-correction ....
Simple over or under correction ....
Perfect flying—no deviation occurred ....

Maximum deviation—altitude .... ft.

Weather during 90° turn:

Smooth ....
Light turbulence ....
Moderate turbulence ....
Heavy turbulence ....

180° turn

a. 180° turn—bank 30°.
b. Airspeed 155 m. p. h.
c. Altitude .... ft. (to be given).

Bank deviation and altitude deviation are each to be observed during specified *fifteen* (15) second intervals in the following sequence:

Instructions Completed	First 15 seconds	Second 15 seconds	Third 15 seconds	Turn Completed
	No measurement	Bank deviation	Altitude deviation	

Maximum deviations are to be recorded for each characteristic for the specified interval *only*.

Heading at beginning of 180° turn ....

Heading at moment wings are first made level ....

Bank deviation (2d 15-second interval—check one):

Failed to correct ....
Oscillating over-correction ....
Simple over or under correction ....
Perfect flying—no deviation occurred ....

Maximum deviation—bank (check one):

± 2° limits ....
± 4° limits ....
± 6° limits ....
Over ± 6° limits ....

Altitude deviation (3d 15-second interval—check one):

Failed to correct ....
Oscillating over-correction ....
Simple over or under correction ....
Perfect flying—no deviation occurred ....

Maximum deviation—altitude .... ft.

Weather during 180° turn:

Smooth ....
Light turbulence ....
Moderate turbulence ....
Heavy turbulence ....

5. *Smooth turn during steep bank:*

a. Bank 45°
b. Initial air speed 165 m. p. h.
c. Minimum air speed 135 m. p. h.
d. Altitude .... ft. (to be given).

Bank deviation and *altitude deviation* are each to be observed during specified ten (10) second intervals in the following sequence:

Instructions Completed	First 10-seconds	Second 10-seconds	Third 10-seconds	Turn Completed
	No measurement	Bank deviation	Altitude deviation	

Maximum deviations are to be recorded for each characteristic for the specified interval only. *Note.*—Consider turbulence and condition of plane when grading.

Bank deviation (2d 10-second interval—check one):

Failed to correct ....
Oscillating over-correction ....
Simple over or under correction ....
Perfect flying—no deviation occurred ....

Maximum deviation—bank (check one):

± 2° limits ....
± 4° limits ....
± 6° limits ....
Over ± 6° limits ....

Altitude deviation (3d 10-second interval—check one):

Failed to correct ....
Oscillating over-correction ....
Simple over or under correction ....
Perfect flying—no deviation occurred ....

Maximum deviation—altitude .... feet.

Degrees to accomplish smooth turn:

Accomplished smooth turn in less than 120° ....
Accomplished smooth turn between 120° and 180° ....
Accomplished smooth turn in over 180° ....
Did not accomplish smooth turn ....
Required by check pilot to roll out of turn because of danger ....

Plane condition—Satisfactory for steep bank: Yes ....; No ....

Weather during steep bank:

Smooth ....
Light turbulence ....
Moderate turbulence ....
Heavy turbulence ....

6. *Stalls—detection and recovery:*

a. Glide without flaps.
b. Power off.
c. Heading .... ° (given by check pilot).
d. Altitude .... ft. (given by check pilot).

Altitude lost in recovery from stall (check one):

0– 50 ft. ....
50–100 ft. ....
100–150 ft. ....
150–200 ft. ....
200–250 ft. ....
250–300 ft. ....
300–350 ft. ....
350–400 ft. ....
400–450 ft. ....
450–500 ft. ....
500–550 ft. ....
550–600 ft. ....
600–650 ft. ....
650–700 ft. ....
Over 700 ft. ....

Maximum deviation—heading:

No deviation ....
± 3° limits ....
± 5° limits ....
Over ± 5° limits ....

Smoothness of control during recovery.—Rate the smoothness of coordination of throttle, rudder, and stick during recovery from stall.

Excellent .... *(10% of pilots)* Good .... *(Majority of pilots)* Poor .... *(10% of pilots)*

Tendency toward second stall:

Excellent procedure—no loss in air speed ....
Faulty procedure—allowed nose to rise reducing air speed ....
Actual tendency toward second stall—slight vibration ....

Weather:

Smooth ....
Light turbulence ....
Moderate turbulence ....
Heavy turbulence ....

7. *Recovery from unusual maneuver.* a. Full panel.

Check Pilot will give definite heading, air speed, and altitude to pilot, then take the controls. He then will perform a maneuver that will place the plane in an unusual position before returning controls to the pilot. Proficiency will be measured in terms of the rate and smoothness with which the pilot regains the initial heading, air speed and altitude. Rating can be given after pilot has corrected heading, and airspeed and has established climb or glide that will regain original altitude.

Heading given was .... °.
Air speed given was .... m. p. h.
Altitude given was .... ft.

Original heading:

Regained quickly ....
Regained average time ....
Regained slowly ....

Original latitude:

Regained quickly ....
Regained average time ....
Regained slowly ....

Original air speed:

Regained quickly ....
Regained average time ....
Regained slowly ....

Smoothness in recovery:

Smooth on controls ....
Rough on controls ....
Jerky on controls ....
Over-controls ....

Assistance during recovery:

No assistance given ....
Assistance by words or signs ....
Assistance by movements of any controls ....

Smoothness of control during recovery. Rate the smoothness of coordination of throttles, rudder, stick during recovery from unusual maneuver.

Excellent .... *(10% of pilots)* Good .... *(Majority of pilots)* Poor .... *(10% of pilots)*

Weather:

Smooth ....

Light turbulence ....

Moderate turbulence ....

Heavy turbulence ....

8. *Power glide.* To right for 1000 feet, straight flight for 30 seconds and then glide to left for 1000 feet, another straight glide and leveling off at a specified altitude and heading.

a. Without flaps.
b. Wheels down.
c. Bank 15°.
d. Air speed 150 m. p. h.
e. Descent 500 ft. per. min.
f. Headings (to be given).

Check Pilot will have pilot put plane on a round numbered altitude, *e.g.*, 3000, 4000, 5000, etc., and will inform pilot of the altitude 1000 feet lower at which the glide is to be completed. At end of right glide and before beginning left glide, pilot is to fly straight for about 30 seconds. Altitude at end of each glide is taken at that moment the wings are made level regardless of gliding attitude of the plane. At end of second glide, check pilot will specify altitude and heading at which pilot is to level off. Check pilot will observe before start of glide the reading of the flight indicator which is indicative of level flight.

Bank deviation and *glide attitude deviation* are each to be observed during specified *fifteen* (15) second intervals in the following sequence:

Instructions Completed	First 15 seconds	Second 15 seconds	Third 15 seconds	Turn Completed
	No measurement	Bank deviation	Glide attitude	

Maximum deviations are to be recorded for each characteristic for the specified interval *only*, using the flight indicator. Deviations in bank are to be measured in degrees from the original bank set up by the pilot. Deviations in glide attitude are to be measured in terms of the linear distance that the horizon line fluctuates from the original position set up by the pilot.

Right turn

Altitude at beginning of right turn .... ft.

Altitude at moment wings are first made level .... ft.

Bank deviation (2d 15-second interval only, check one):

Failed to correct ....
Oscillating over-correction ....
Simple over or under correction ....
Perfect flying—no deviation ....

Maximum deviation—bank (check one):

± 2° limits ....
± 4° limits ....
± 6° limits ....
Over ± 6° limits ....

Glide attitude deviation (3d 15-second interval only):

Failed to correct ....
Oscillating over-correction ....
Simple over or under correction ....
Perfect flying—no deviation ....

Maximum deviation—glide attitude (check one):

± $\frac{1}{16}$-inch limits ....
± $\frac{1}{8}$-inch limits ....
± $\frac{1}{4}$-inch limits ....
± $\frac{1}{2}$-inch limits ....

Left turn

Altitude at beginning of left turn .... feet.

Altitude at moment wings are first made level .... feet.

Bank deviation (2d 15-second interval only, check one):

Failed to correct ....
Oscillating over-correction ....
Simple over or under correction ....
Perfect flying—no deviation ....

Maximum deviation—bank (check one):

± 2° limits ....
± 4° limits ....
± 6° limits ....
Over ± 6° limits ....

Glide attitude deviation (3d 15-second interval only, check one):
- Failed to correct ....
- Oscillating over-correction ....
- Simple over or under correction ....
- Perfect flying—no deviation ....

Maximum deviation—glide attitude (check one):
- ± 1/16-inch limits ....
- ± ⅛-inch limits ....
- ± ¼-inch limits ....
- ± ½-inch limits ....

Straight glide

Specified heading for leveling off .... °.
Actual heading at which leveled off .... °.
Specified altitude for leveling off .... feet.
Actual altitude at which leveled off .... feet.
Trimmed plane (for whole maneuver): Yes ....; No ....

Weather during glide:
- Smooth ....
- Light turbulence ....
- Moderate turbulence ....
- Heavy turbulence ....

APPENDIX F

Airplane Commander Rating Scales

Appendix F.1

RATING OF AIRPLANE COMMANDERS

Scale B

.... (Name of rater)

.... (Date)

.... (Duty assignment)

Numbers of weeks experience with this group of students ....

Instructions to raters

You are being asked to rate each of several Airplane Commanders of crews in training at your station. For each of the items on this rating scale, you should read the question at the top and each of the five statements under it. Then for each student, decide which one of the five statements in *your opinion* best describes that student and his abilities and probable reactions. Then, beside the name of that student, put an "X" in the box having the same column number as the statement you have selected. It is doubtful that any statement will describe a student exactly, but you are being asked to use your judgment in deciding which one of the five statements *most nearly* fits each student. Avoid using the opinion of others; this rating should be *your* opinion of the Airplane Commanders listed.

Item A. How strongly would this airplane commander be recommended for a possible lead crew assignment?

1. This airplane commander has done an outstanding job and definitely would be recommended as a lead airplane commander.
2. This airplane commander has done well and would be strongly considered as lead crew material.
3. This airplane commander has done what was expected of him and might be recommended for a lead crew position provided he works harder.
4. This airplane commander probably would not be recommended for a lead assignment although he has completed the requirements.
5. This airplane commander has scarcely completed the minimum essentials and would not be considered as lead calibre.

1	2	3	4	5

Item B. How well has this airplane commander mastered formation flying?

1. This airplane commander flies very well in all formation positions and operates smoothly with excellent judgment in anticipating actions of others.
2. This airplane commander is good in formation flying, observes standing operating procedures, and easily maintains his position.
3. This airplane commander understands formation flying and is fair in assembly and break-up of formations.
4. This airplane commander maintains his proper position with some difficulty and is a little rough in formation penetration and reassembly.
5. This airplane commander sometimes fails to follow standing operating procedures and occasionally neglects to warn other crews of flight changes.

1	2	3	4	5

Item C. How eager and interested is this airplane commander in his job and training activities?

1. This airplane commander is an eager beaver, works hard at extra tasks, learns new techniques, and completes training ahead of others.
2. This airplane commander likes his job and shows more than average interest by putting in extra hours on trainers, missions, and by going to classes of other crew members.
3. This airplane commander does all that is expected of him and shows average interest in the progress of his crew.
4. This airplane commander seems to display a lack of interest but usually can be relied upon to complete the minimum essentials of training.
5. This airplane commander seems to be disinterested in his job, slovenly in accomplishing his training, and unconcerned about improving himself.

1	2	3	4	5

Item D. How well does this airplane commander prepare, plan ahead, and coordinate the activities of his crew?

1. This airplane commander always plans ahead, uses excellent judgment, and never leaves anything to chance for which he can prepare.
2. This airplane commander usually plans the activities of his crew in some detail, uses good judgment, and seldom is hurried in making decisions.
3. This airplane commander plans and prepares adequately, uses fair judgment, but sometimes is hesitant about effecting appropriate action.
4. This airplane commander plans the flight activities of his crew fairly well but sometimes neglects important factors and has some difficulty in making necessary decisions.
5. This airplane commander seldom prepares sufficiently well, seems to become confused under pressure, and makes mistakes.

1	2	3	4	5

Item E. How adequate is this airplane commander's ability to lead and supervise?

1. This airplane commander obtains the maximum from his own team and has the respect of other crews. He is exceptional.
2. This airplane commander is conscientious, understanding, and reasonably skillful as a leader.
3. This airplane commander accepts his responsibilities and shows average effort in leading the activities of his crew.
4. This airplane commander knows what is expected of him but fails to obtain the confidence of his crew.
5. This airplane commander lacks understanding of his responsibilities and thinks his job is done when he lands the aircraft.

1	2	3	4	5

Item F. How skillful is this B-29 airplane commander in instrument flying?

1. This airplane commander flies on instruments with precision, concentrates with ease, and does not become confused.
2. This airplane commander is more skillful than the average in flying by instruments and seldom becomes fatigued.
3. This airplane commander demonstrates average proficiency in instrument flying and tires no more than may be expected.
4. This airplane commander seems to be weak in radio orientation and instrument let-down procedures and sometimes misinterprets his instruments.
5. This airplane commander is somewhat deficient in instrument flying techniques and seems to lack adequate understanding of aircraft instruments.

1	2	3	4	5

Item G. How would you rank this airplane commander in over-all efficiency and ability among all the airplane commanders now in training at your station?

1. In the highest 10 percent—one of the very best.
2. In the upper 20 percent—very good.
3. In the middle 40 percent—competent and able.
4. In the lower 20 percent—fairly proficient.
5. In the lowest 10 percent—minimum satisfactory.

1	2	3	4	5

Appendix F.2

RATING OF AIRPLANE COMMANDERS
Scale A

.... (*Name of rater*)

.... (*Date*)

.... (*Position*)

Instructions to raters

You are being asked to rate each of several airplane commanders on crews in training at your station. For each of the items on this rating scale you should read the question at the top and each of the five statements under it. Then for each student, decide which one of the five statements in *your* opinion best describes that student or his abilities and probable reactions. Then place a check mark in the box having the same column number as the number of the statement which best describes each one and which is immediately opposite the appropriate crew number. It is doubtful that any statement will exactly describe a student, but you are being asked to use your judgment in deciding which one of the five statements *most nearly* fits each student. Avoid using the opinion of others; this rating should be *your* opinion of the airplane commanders listed.

Some airplane commanders appear to be able to anticipate impending events more rapidly than others. This may be called foresight. For example, anticipation of turns, changes in altitude, gasoline consumption, necessary equipment, flak, I.P., weather, etc. How much foresight does this airplane commander demonstrate?

1. This airplane commander has the unique ability to foresee almost every impending occurrence and their complete implications. He invariably evades their undesirable effects by appropriate action.
2. This airplane commander is always alert to the possibilities of impending occurrences and their probable implications. He minimizes their effect by wise preventive and corrective action.
3. This airplane commander is generally alert to the possibilities of impending occurrences and their probable implications. He is generally adept in initiating the proper preventive and corrective action.
4. This airplane commander sometimes sees the possibility of some situations occurring. He may fail to see all the implications involved and occasionally has difficulty in choosing the proper corrective action.
5. This airplane commander frequently fails to see the possibility of situations occurring. He often fails to see all implications involved and is sometimes confused with regard to the proper preventive procedure.

Crew Number	Name	ASN	5	4	3	2	1

How much interest does this airplane commander take in his crew?

1. This airplane commander is enthusiastically interested in all phases of the crew activities both personal and technical. He is constantly alert to assist them in any reasonable manner even if it results in embarrassment for himself.

2. This airplane commander is interested in all phases of crew activity both personal and technical. He is usually alert to assist them in any reasonable manner even if it results in added effort on his part.

3. This airplane commander is usually interested in all phases of crew activity. He will always help them with their personal and technical problems.

4. This airplane commander rarely evidences interest in the activities of the crew. However, when crew members present their personal or technical problems to him he usually helps solve them.

5. This airplane commander is obviously disinterested in his crew. He apparently is not interested in their personal problems and is unconcerned with their technical improvement.

Crew Number	Name	ASN	5	4	3	2	1

APPENDIX G

Check List of Airplane Commander Ability

Appendix G.1

CHECK LIST OF AIRPLANE COMMANDER ABILITY

INFORMATION:

Name .... Rank .... ASN ....

Crew number .... Weeks in training ....

Type of training mission ....

Type of aircraft .... Hours this aircraft ....

INFORMATION ON CHECK PILOT:

Name .... Rank .... ASN ....

Date .... Field ....

	Good or what top 10% of pilots would do	Average or what average pilots would do	Poor or what lowest 10% of pilots would do	Failed to make check
Before take-off checks				
....	....	....	....	....
1. Visual outside check (manner in which check is conducted)..	....	....	....	....
2. Visual interior check (manner in which check is conducted)..	....	....	....	....
3. Check on crew members preflighting of positions	....	....	....	....
4. Personal equipment inspection	....	....	....	....
5. Briefing of crew	....	....	....	....
6. Instruction to crew (if necessary)	....	....	....	....
7. Cockpit procedure and check	....	....	....	....
8. Pilot's manner of making all checks (e.g., systematic, business-like)	....	....	....	....
9. Taxi procedure	....	....	....	....
10. Check if crew is in take-off position: Yes ....; No ....:				

Remarks:

During flight:

	Yes	No
1. Ordered crew to positions after making 500' turn	......	......
2. Called for crew position report after giving crew position report	......	......
3. Reporting planes sighted by crew	......	......
4. Pilot has crew members change positions during flight	......	......
5. Pilot keeps crew informed as to progress of mission	......	......
6. Instruction or explanation to crew during flight	......	......
7. One crew member in pilot's compartment on interphone at all times	......	......
8. Emergency procedures	......	......
9. Frequent checks on crew positions	......	......
10. Allows crew to sleep during mission	......	......

	Good or what top 10% of pilots would do	Average or what average pilots would do	Poor or what lowest 10% of pilots would do	Failed to make check
11. Interphone procedure	....	....	....	....
12. Orderliness of plane	....	....	....	....
13. Crew cooperation ..	....	....	....	....

14. Calls crew for landing:

Yes ......; No ....

Remarks:

After landing:

	Yes	No
1. Makes after landing checks	......	......
2. Fills out required forms	......	......
3. Checks crew members, critiques mission	......	......
4. Checks errors, gives instruction	......	......
5. Checks police of plane and ramp	......	......

Remarks:

General:

	Good or what top 10 per-cent of pilots would do	Average or what average pilots would do	Poor or what lowest 10 per cent of pilots would do	Failed to make check
1. Overall rating as an airplane commander..	....	....	....	....
2. General pilot ability..	....	....	....	....
3. General pilot-crew relations	....	....	....	....
4. Utilization of "waiting time"	....	....	....	....
5. Ground discipline	....	....	....	....
6. Air discipline	....	....	....	....

7. Called officers by correct titles

Yes ....; No ....

8. Called enlisted men by correct titles

Yes ....; No ....

Remarks:

APPENDIX H

AAF Letter 50-117. Subject: Screening of Combat Crew Personnel

Appendix H.1

50-117
2 Pages
Page 1

HEADQUARTERS, ARMY AIR FORCES,
Washington, 7 June 1945.

AAF Letter 50-117.
SUBJECT: Screening of Combat Crew Personnel.
To: Commanding Generals, Major AAF Commands.
Director, Air Technical Service Command.
Commanding Generals and Commanding Officers, Independent AAF Activities.
Commanding Generals and Commanding Officers, Subordinate AAF Commands and AAF Divisions, Wings and Districts.
Commanding Officers, AAF Base Units (not included above).

(This Letter supersedes AAF Letter 50-117, 1 May 1945.)

1. One of the major problems in the theaters is the selecting and training of qualified lead crews. At present the burden of forming lead crews has been placed on the tactical organizations in the field after the crews have completed their training in the Zone of the Interior. In the process of forming lead crews in the tactical organizations it is quite often necessary to break up two or three crews, thus voiding much of the crew training and breaking up the integrity of the combat crew. It is believed that the following procedures will facilitate the identification and training of lead crews and reduce the number of necessary changes in crew personnel.

2. The assigment of personnel to crews and the designation of certain crews as potential lead crews will be made at the AAF Combat Crew Processing and Distribution Center in the AAF Training Command on the basis of a lead crew aptitude score given to each officer aircrew member. This lead crew aptitude score will be a weighted average of evaluations of aerial training and experience, written proficiency examinations in appropriate specialties, and original aircrew aptitude test scores.

3. The lead crew aptitude score for each crew member will be entered on his AAF Individual Training Record Form as provided for in AAF Letter 50-55. If he is assigned to a potential lead crew the letter "L" will be entered after the lead crew aptitude score. If during subsequent training he is removed from a potential lead crew for lack of proficiency, suitable notation will be made in the "Remarks" section.

4. Potential lead crews will be composed of aircrew members who have the highest lead crew aptitude scores, namely, seven, eight, and nine. Insofar as possible, potential lead crews will be made up of crew members all of whom have lead crew aptitude scores of nine, all of whom have lead crew aptitude scores of eight, or all of whom have lead crew aptitude scores of seven. The remaining combat crews will be assembled without regard to lead crew aptitude scores.

5. A Combat Crew Record Form (copy attached) will be initiated at the AAF Combat Crew Processing and Distribution Center, which will indicate whether the crew is designated as a potential lead crew or as a combat crew. The name, rank, and serial number of each crew member assigned to the crew will be entered. This form will be forwarded in duplicate to the combat crew training station to which the crew is sent. Authority is granted to the AAF Training Command for the reproduction of this form as shown in the attachment to this Letter, in sufficient quantities to furnish a three (3) months supply for use at the AAF Combat Crew Processing and Distribution Center.

6. Upon completion of combat crew training, the designation of potential lead crew or of combat crew made in the AAF Training Command will be confirmed or revised at the training air force station and suitable notations will be made on the form. Appropriate remarks concerning the proficiency of the crew as a whole will be entered.

7. The original of the Combat Crew Record Form will be forwarded to the theater for the information of the combat organization to which the crew is assigned. The duplicate will be forwarded to the commanding general of the air force in which combat crew training is accomplished. The surgeon and the office responsible for training programs at the headquarters of the air force will use this information to evaluate the potential lead crew selection program.

8. If, at any time subsequent to original assembly of the crew, it may become necessary to substitute for any crew member, every effort will be made to replace him with an aircrew member having the same lead crew aptitude score.

9. A clearance number is not required for the Combat Crew Record Form under provisions of paragraph 4a (2), AAF Regulation 80-2.

By Command of General Arnold:

[SEAL]

Ira C. Eaker,
Lieutenant General, United States Army,
Deputy Commander, Army Air Forces.

Attachment: Sample Combat Crew Record Form.
Distribution: Air forces and AAF commands overseas.

COMBAT CREW RECORD FORM

AAF TRAINING COMMAND

Section I:

Airplane commander (Last name first)	Date	Type of aircraft	Crew designation	Crew seat No:
1.				

2. Potential lead crew ☐ Combat crew ☐

Crew Roster

Last name First name MI Grade ASN	Last name First name MI Grade ASN
Co-Pilot 3.	Gunner
Navigator 4.	Gunner
Bombardier 5.	Gunner
Flight engineer 6.	Gunner
Radar observer 7.	Gunner
Gunner 8.	

Remarks 9.	Authenticated by: Station:

Section II:

...... *Air Force* (Training)

Training station:	Class	Date	Average circular error of crew
10.			

11. Potential lead crew ☐ Combat crew ☐

Record of Crew Changes

Position	Last name	First name	MI	Grade	ASN	Reason for removal of original Member
12.						
13.						
14.						
15.						

Remarks on crew performance

16.

Authenticated by:

INSTRUCTIONS FOR COMPLETION OF COMBAT CREW RECORD FORM

The Combat Crew Record Form is designed, in accordance with AAF Letter 50-117, to provide a continuous evaluative record of the bombardment crew from the time it is assembled until it completes CCTS training. Section I is to be completed in the AAF Training Command and Section II in the training air force. Specific directions for completion follow:

Section I, AAF Training Command. The form will be initiated in duplicate at the AAF Combat Crew Processing and Distribution Center and both copies will be mailed directly to the commanding officer of the combat crew training station to which the crew is sent.

Line 1

a. Name, grade, and ASN of airplane commander at the time of crew assignment. If the airplane commander is subsequently replaced, a line will be drawn through this entry and the new airplane commander's name will be entered in the appropriate space below.

b. Type of aircraft for which the crew is assembled.

c. Code number or current symbol for designation of the crew.

d. Name of the CCTS to which the crew is sent for training.

Line 2

On the basis of procedures outlined in paragraph 2 of the AAF Letter cited above and specific directives from the Commanding General, AAF Training Command, the crew will be evaluated as potential lead crew or combat crew and an "X" will be placed in the appropriate box.

Line 3 through 8

Using only spaces which are applicable for the type of aircraft to which the crew is assigned, names, grades, and ASN's of crew members will be entered. When crew changes are made a line will be drawn through the name of each member replaced and new members will be listed as specified below.

Line 9

Remarks on any additional pertinent information regarding the assembly of a particular crew. The officer in charge of crew assembly will authenticate this section.

Section II, Training Air Force. Two copies of the form will be completed. The original will be placed in a sealed envelope and entrusted to the airplane commander for delivery to the commanding officer of the combat group to which the crew is assigned in the theater of operations. The duplicate will be for-

warded to the Commanding General of the air force in which CCTS training, attention surgeon, is accomplished.

Line 10

a. Name of CCTS.
b. Class in which the crew completes training.
c. Date of completion of Section II.
d. Average circular error of the crew made in completing the minimum requirement for bombs dropped as outlined in the AAF Training Standards 20-1, 20-2, or 20-3.

Line 11

On the basis of performance in CCTS the crew will be designated as a potential lead crew or a combat crew and an "X" placed in the appropriate box.

Lines 12 through 15

Crew changes will be recorded by indicating the crew position of the new member (airplane commander, navigator, etc.), name, grade and ASN of the new member, and the reason for removal of the original member (medical, furlough, lack of proficiency, etc.).

Line 16

Additional evaluation of the crew will be entered under "Remarks." This section will be authenticated by the officer in charge of training.

APPENDIX L

Experience and Opinions Questionnaire

Appendix L1

OFFICER QUESTIONNAIRE

(Based on 528 cases unless otherwise indicated)

This questionnaire is to determine and record your experiences and opinions as an officer returned from combat. Army Air Forces are vitally interested in finding out your reaction to your training, to what you have done and seen while in combat and to your present assignment. This questionnaire is not to be considered a test of any kind but rather as a record of what you have seen, what you have done and what you think.

You will find that the questions may be answered by indicating the number of one of the alternative answers in the space provided on the answer sheet. Be sure that you write the number indicating your answer in the space corresponding to the question. However, in some of the questions you will find that the alternatives listed do not include your answer. Instructions are given with these questions to write your answer in the space provided on the *back* of your answer sheet. Read the question and all of the alternative answers before answering the question. Do not skip any question. Answer each one to the best of your ability.

This booklet of questions will be used again for another returned officer, therefore, you are requested not to make any pencil marks upon it or deface it in any way. Write your name, serial number, present post or field and date on the answer sheet in the space provided. Record the principal MOS held in combat and all other MOS's in which you are qualified on the lines provided. Your name is desired in order that further information about you may be obtained from your Form 66-2. Your name will not accompany any report of the findings of this study. Do not hesitate to choose the answers that express the way you really feel and think. There will be no time limit.

The first question asks, "What is your present grade?" Under the question is a list of grades from Colonel to Flight Officer. Pick out your present grade and write the number corresponding to it on the list of alternative answers in the space under "1" on your answer sheet. Continue in the same manner for the remainder of the questions.

OFFICER QUESTIONNAIRE

Age	Percent	Age	Percent
21 or less	1.9	26	17.2
22	3.4	27	11.2
23	11.5	28	9.3
24	17.2	29	6.1
25	18.0	30 or more	4.2

(Percentage distribution of answers is indicated in margin.)

(1) What is your present grade? (Write the correct number in space No. 1)

0.0 1 Colonel.
.2 2 Lieutenant Colonel.
5.2 3 Major.
34.3 4 Captain.
52.5 5 First Lieutenant.
4.7 6 Second Lieutenant.
2.6 7 Flight Officer.
8 Other (indicate what on back of answer sheet next to No. 1).

(2) What is your age in years, to your nearest birthday? (Write it in space No. 2)

(3) Are you married or single? (Write only one number in space No. 3)

38.1 1 single.
.0 2 widowed.
.9 3 divorced or separated other than by military necessity.
33.9 4 married, *before* going overseas.
.9 5 married, *while* overseas.
26.1 6 married, *since* returning from overseas.

(4) How many dependents do you have? (Include wife, dependent parents or children)

40.0 1 one.
20.9 2 two.
1.1 3 three.
.6 4 four.
.6 5 five or more.
36.0 6 none.

(5) Before entering the army how far did you go in school?

27.7 1 high school and/or vocational school graduate or less.
10.0 2 one year college.
26.3 3 two years college.
16.1 4 three years college.
14.0 5 four years college or graduated from college.
2.8 6 did post graduate work in college or university.
2.5 7 attended or graduated from professional school (dental, medical, engineering, etc.).
.4 8 attended United States Military Academy but did not graduate.
.2 9 United States Military Academy graduate.

(6) What was your principal duty or aeronautical rating during MOST of the time you were in combat theaters? (Write only one number in space No. 6.)

54.9 1 Pilot.
4.4 2 Copilot.
14.6 3 Bombardier.
19.1 4 Navigator.
5.9 5 Bombardier-Navigator.
.4 6 Aircraft Observer.
.0 7 Officer: Flight Engineer.
.0 8 Officer: Radar Observer.
.8 9 Other (write what on the back of the answer sheet after No. 6).

(7) Did you fly a lead position in a flight, squadron or group formation for at least five missions? If you held both a squadron and a group position record only the group position. If you held a squadron or group position not included in the list below, write in alternative 10 and write the name of the position on the back of the answer sheet next to No. 7. (Write only one number in space No. 7.)

8.1 1 Lead pilot, group
16.7 2 Lead pilot, squadron.
18.7 3 Lead pilot, flight.
5.7 4 Lead bombardier, group.
6.8 5 Lead bombardier, squadron.
3.8 6 Lead bombardier, flight.
5.9 7 Lead navigator, group.
8.3 8 Lead navigator, squadron.
2.8 9 Lead navigator, flight.
23.1 10 other (Write what on back of answer sheet after No. 7) or did not hold group, squadron or flight position.

Various expressions of opinion concerning the motivation of men in combat have been given. It is known that the motivating factors are different during the initial missions than during the remainder of the tour of combat. Indicate the degree to which each of the following factors motivated you during the *second half* of your stay in the combat theater.

(8) Thought it was the best way to strike an active blow against the enemy.
40.5 1 influenced me greatly.
48.7 2 influenced me somewhat.
10.0 3 did not influence me.

(9) Desired to do a good job and contribute my share.
78.4 1 influenced me greatly.
19.8 2 influenced me somewhat.
.9 3 did not influence me.

(10) Didn't want to let the other members of the crew down by quitting.
49.1 1 influenced me greatly.
25.4 2 influenced me somewhat.
24.2 3 did not influence me.

(11) Liked the excitement of combat.
18.0 1 influenced me greatly.
42.2 2 influenced me somewhat.
38.6 3 did not influence me.

(12) Preferred flying duty to ground duty.
71.2 1 influenced me greatly.
19.7 2 influenced me somewhat.
8.3 3 did not influence me.

(13) Performed the duties only because I was ordered to do so.
3.4 1 influenced me greatly.
18.0 2 influenced me somewhat.
77.1 3 did not influence me.

Judging from your experience in combat from what age periods do the *most* successful combat officers come? Make your judgment on the basis of the performance of the officer *in combat*. There are four questions to be answered: Question 14 for bomber pilots and co-pilots. Question 15 for fighter pilots. Question 16 for bombardiers and question 17 for navigators.

(14) Age period of most successful combat bomber pilots and bomber copilots. (Based on 335 cases, 36.6 percent of the 528 felt they were not qualified to answer.)

.0	1	18–19
3.3	2	20–21
28.3	3	22–23
52.6	4	24–25
14.1	5	26–27
1.2	6	28–29
.3	7	30–31
.0	8	32–33
.3	9	34 or over

(15) Age period of most successful combat fighter pilot. (Based on 364 cases, 31.0 percent of the 528 felt they were not qualified to answer.)

1.1	1	18–19
26.4	2	20–21
50.0	3	22–23
18.4	4	24–25
3.0	5	26–27
1.1	6	28–29
	7	30–31
	8	32–33
	9	34 or over

(16) Age period of most successful combat bombardiers. (Based on 309 cases, 41.5 percent of the 528 felt they were not qualified to answer.)

0.0	1	18–19.
7.4	2	20–21.
33.5	3	22–23.
39.5	4	24–25.
11.9	5	26–27.
2.6	6	28–29.
	7	30–31.
	8	32–33.
	9	34 or over.

(17) Age period of most successful combat navigators. (Based on 319 cases, 39.6 percent of the 528 felt they were not qualified to answer.)

0.0	1	18–19.
2.8	2	20–21.
27.0	3	22–23.
47.6	4	24–25.
19.4	5	26–27.
2.2	6	28–29.
.6	7	30–31.
.3	8	32–33.
.0	9	34 or over.

(18) How long were you overseas?

2.1	1	4 months or less.
4.2	2	5 or 6 months.
8.7	3	7 or 8 months.
32.8	4	9, 10, or 11 months.
24.4	5	12, 13, or 14 months.
15.7	6	15, 16, or 17 months.
5.1	7	18, 19, or 20 months.
3.0	8	21, 22, or 23 months.
4.0	9	24 or more months.

(19) How many missions did you complete outside of the continental United States? (By authority of the Intelligence Officer all men are permitted to answer this question.)

4.2 1 10 or less.
3.6 2 11 to 20.
8.1 3 21 to 25.
6.8 4 26 to 30.
15.3 5 31 to 40.
20.3 6 41 to 50.
16.5 7 51 to 60.
6.4 8 61 to 70.
17.4 9 more than 70.

(20) How many combat flying hours did you complete? (By authority of the Intelligence Officer, all men are permitted to answer this question.)

11.0 1 below 100 hours.
42.4 2 100 to 200 hours.
30.8 3 200 to 300 hours.
11.6 4 300 to 400 hours.
1.1 5 400 to 500 hours.
1.1 6 500 to 600 hours.
.4 7 600 to 700 hours.
.4 8 700 to 800 hours.
1.3 9 800 hours or more.

(21) Did you fly more or less than the regularly required number of missions or hours for your theater and type of plane? (Write only one number in space No. 21)

9.8 1 flew less than the required number of missions or hours.
29.2 2 flew exactly the required number of missions or hours.
19.5 3 flew more than the required number of missions or hours.
41.3 4 there was no exact policy on the number of missions or hours required or do not know.

(22) In what theater did you fly all or *most* of your combat missions?

16.7 1 European—(England-France).
44.3 2 Mediterranean (Italy, North Africa, Middle East).
27.1 3 Pacific (Central, South, Southwest).
4.0 4 China-Burma-India.
2.3 5 Caribbean.
4.0 6 Alaska.
.4 7 North Atlantic—Greenland.
1.3 8 other (if you flew in a theater not listed here, write its name on the back of the answer sheet after No. 22).

(23) In what *one* type of plane did you fly all or most of your combat missions? (Write only one number in space No. 23.)

18.2 1 B-17.
3.8 2 B-24.
32.4 3 B-25.
16.9 4 B-26.
2.7 5 A-20.
8.7 6 P-40.
3.8 7 P-51.
4.2 8 P-39.
1.9 9 P-38.
7.6 10 other.

(If you flew a type not listed here, write in alternative 10, and write its type or name on the back of the answer sheet after No. 23.)

(24) How many crash landings, forced landings, or ditchings *in friendly territory* have you been in? (Write the actual number in space No. 24 on your answer sheet.)

(25) How many crash landings, forced landings, or ditchings *in enemy territory or territory not controlled by friendly troops* have you been in? (Write the actual number in space No. 25 on your answer sheet.) (By authority of the Intelligence Officer all returnees may answer this question.)

(26) On what proportion of your missions did you meet enemy opposition of some kind? Opposition includes flak, fighters, etc. (Write only one number in space No. 26.)

4.0 1 never.
4.5 2 rarely.
5.9 3 less than half of the missions.
16.7 4 half or more of the missions.
52.1 5 almost every mission.
16.7 6 every mission.

(27) Did you ever "bail out" on a combat flight? (Write only one number in space No. 27.)

4.0 1 yes; over friendly territory.
2.3 2 yes; over enemy territory, or territory not controlled by friendly troops (by authority of the Intelligence Officer, returnees may give this answer if it is true).
.0 3 yes; over both friendly and enemy territory at different times.
93.6 4 no; never bailed out.

(28) Were you ever wounded in combat? (Write only one number in space No. 28. If you were wounded more than once, answer for the worst one.)

10.0 1 slight wound; not hospitalized.
4.0 2 slight wound; in hospital less than one week.
.6 3 moderately severe wound but not hospitalized.
2.3 4 moderately severe wound; in hospital up to 6 weeks.
2.1 5 severe wound; in hospital more than 6 weeks.
80.7 6 was never wounded.

(29) While on foreign service, were you ever *injured* in connection with aircraft while *not* in combat? (Write only one number in space No. 29. If you were injured more than once, answer for the worst injury.)

6.6 1 slight injury; not hospitalized.
1.3 2 slight injury; in hospital less than one week.
.4 3 moderately severe injury; not in hospital.
1.7 4 moderately severe injury; in hospital up to 6 weeks.
1.7 5 severe injury; in hospital more than 6 weeks.
87.5 6 was never injured.

(30) How much were you bothered by air sickness while in combat?

93.6 1 never.
5.9 2 only on a few flights.
.4 3 frequently.
.2 4 on nearly every flight.
.0 5 on every flight.

(31) While on foreign service, were you ever—(Mark the number of the statement that applies to you. Write only one number in space No. 31.)

4.4 1 sent to a hospital because of operational fatigue or exhaustion.
13.6 2 sent to a rest camp because of operational fatigue.
6.6 3 given leave(s) for reason of operational fatigue.
4.0 4 diagnosed as having flying fatigue or operational fatigue, but not treated for it.
24.6 5 sent to a rest camp routinely, not for operational fatigue.
19.5 6 never had operational fatigue nor sent to a rest camp.
27.1 7 given routine combat leave.

(32) Were any of the members of the crew killed in the planes in which you flew in combat? (Write only one number in space No. 32.)

7.4 1 1 was killed.
2.5 2 2 were killed.
.9 3 3 were killed.
.0 4 4 were killed.
.0 5 5 were killed.
.6 6 6 were killed.
.6 7 7 were killed.
.8 8 8 or more were killed.
72.5 9 none was killed.

(33) Were any of the members of the crew wounded in the planes in which you flew in combat? (Write only one number in space No. 33.)

23.1 1 1 to 2 were wounded.
12.1 2 3 to 4 were wounded.
3.6 3 5 to 6 were wounded.
1.9 4 7 to 8 were wounded.
.6 5 9 to 10 were wounded.
.0 6 11 to 12 were wounded.
.2 7 13 to 14 were wounded.
.4 8 15 or more were wounded.
43.6 9 none was wounded.

(34) How many of your close friends, *not members of your crew*, were missing, killed or wounded? (Those who flew in the same plane are not to be considered. Write only one number in space No. 34.)

17.2 1 half or more of them were missing, killed or wounded.
27.7 2 between a quarter and a half were missing, killed or wounded.
34.7 3 several were missing, killed or wounded, but less than a quarter.
15.0 4 only a few were missing, killed or wounded.
2.5 5 no close friend was missing, killed or wounded.

(35) What was the *principal* reason for your being returned to the United States from an overseas theater? (Write only one number in space No. 35.)

55.1 1 rotation policy; having flown the required number of missions or hours.
18.5 2 returned because of flying fatigue, after completing about the usual number of missions in my theater.
2.3 3 returned because of flying fatigue or operational fatigue without completing the usual number of missions.
3.4 4 wounded in action or injured in connection with aircraft.
3.6 5 illness, injury or operation not connected with aircraft.
1.3 6 escaped prisoner of war or evader of capture (On instructions of Intelligence officer returnees may give this reason if it is true.)
5.[illegible] 7 returned for instruction or reassignment.
.8 8 personal request to be returned.
9.5 9 other (Indicate what on back of answer sheet after No. 35.)

(36) After you returned from overseas duty, were you processed in a Redistribution Center between 24 May 1944 and 6 August 1944? (The three Redistribution Centers are at Atlantic City, N.J., Miami Beach, Fla., and Santa Monica, Calif.)

25.9 1 not processed in Redistribution Center.
5.9 2 processed at Atlantic City between 24 May 1944 and 6 August 1944.
8.3 3 processed at Miami Beach between 24 May 1944 and 6 August 1944.
4.7 4 processed at Santa Monica between 24 May 1944 and 6 August 1944.
54.7 5 processed at Atlantic City, Miami Beach or Santa Monica but not during period 24 May 1944 to 6 August 1944.

(37) How long have you been in this country since you returned from duty overseas?

3.8 1 2 months or less.
12.9 2 3 or 4 months.
13.1 3 5 or 6 months.
17.4 4 7 or 8 months.
13.1 5 9 or 10 months.
13.4 6 11 or 12 months.
9.1 7 13 or 14 months.
7.6 8 15 or 16 months.
9.7 9 17 months or more.

(38) Which one of the following conditions in the United States are you *most* dissatisfied with? (Write only one number in space No. 38.)

44.1 1 the attitude of labor in this country.
8.7 2 the attitude of business and industrial management in this country.
11.4 3 the general attitude of civilians to military personnel.
12.9 4 the nature of military courtesy, discipline and training in this country.
1.3 5 rationing.
18.1 6 the attitude of military personnel who have never been overseas.
.4 7 my own family relationships.

(39) Which one of the following types of assignment do you have at present?

65.2 1 assignment involving flying duties as an instructor.
16.9 2 assignment involving flying duties not as an instructor.
5.3 3 assignment involving ground duties as an instructor.
12.7 4 assignment involving ground duties of an administrative nature (not as an instructor).

(40) Which one of the following phrases *best* describes your satisfaction with your present assignment?

27.3 1 very satisfactory.
44.7 2 satisfactory.
10.6 3 indifferent.
13.3 4 unsatisfactory.
4.0 5 very unsatisfactory.

(41) In your opinion, how well is the Army now making use of your ability and experience? (Include either or both military and civilian experience.)

20.8 1 best use of my ability and experience.
55.5 2 good use of my ability and experience.
19.9 3 very little use of my ability and experience.
3.6 4 no use of my ability and experience.
.2 5 have no present assignment.

(42) Regardless of what you are doing at the present time, which one of the following types of assignment appeals to you most?

33.1 1 assignment involving flying duties as an instructor.
52.7 2 assignment involving flying duties not involving instructing.
4.4 3 assignment involving ground duties as an instructor.
9.8 4 assignment involving ground duties of an administrative nature (not as an instructor).

(43) Regardless of what you are doing at the present time, which one of the following best expresses your desire relative to assignment to flying duties as an instructor?

33.7 1 desire assignment to flying duties as an instructor.
33.3 2 indifferent to assignment to flying duties as an instructor.
33.0 3 desire not to be assigned to flying duties as an instructor.

(44) Regardless of what you are doing at the present time, which one of the following best expresses your desire relative to assignment to flying duties not involving instructing?

51.3 1 desire assignment to flying duties *not* involving instructing.

32.4 2 indifferent to assignment to flying duties *not* involving instructing.

15.9 3 desire not to be assigned to flying duties not involving instructing.

(45) Regardless of what you are doing at the present time, which one of the following best expresses your desire relative to assignment to ground duties as an instructor?

8.1 1 desire assignment to ground duties as an instructor.

16.1 2 indifferent to assignment to ground duties as an instructor.

75.8 3 desire not to be assigned to ground duties as an instructor.

(46) Regardless of what you are doing at the present time, which one of the following best expresses your desire relative to assignment to ground duties of an administrative nature (not as an instructor).

20.5 1 desire assignment to ground duties of an administrative nature (not instructor).

24.2 2 indifferent to assignment to ground duties of an administrative nature (not instructor).

55.3 3 desire not to be assigned to ground duties of an administrative nature (not instructor).

(47) How strong is your desire to remain on flying status?

74.1 1 very strong.

15.7 2 strong.

4.9 3 indifferent.

1.9 4 somewhat opposed to it.

.4 5 strongly opposed to it.

3.0 6 not on flying status at present time.

(48) At present, how much are you bothered by air sickness?

90.2 1 never.

6.3 2 only on a few flights.

1.3 3 frequently.

.0 4 on nearly every flight.

.0 5 on every flight.

2.3 6 not on flying status.

(49) Are you on limited service at the present time?

94.7 1 no..

2.5 2 yes; physical ailment.

2.7 3 yes; operational fatigue or combat fatigue.

.2 4 yes; other non-physical reason not mentioned in 3 above.

It has been suggested that the following factors influence officers' feelings relative to returning to combat duty. Indicate the degree to which each of the following factors influences your feelings *toward* returning to combat.

(50) *Want* to return to square accounts because of my friends who have been wounded or killed overseas.

1.3 1 influences me greatly.

21.6 2 influences me somewhat.

77.1 3 does not influence me.

(51) *Want* to return because I prefer duty in combat to the duty to which I am now assigned.

11.0 1 influences me greatly.

24.1 2 influences me somewhat.

65.0 3 does not influence me.

(52) *Want* to return because I am dissatisfied with conditions as I have found them here in the United States.

9.3 1 influences me greatly.
23.9 2 influences me somewhat.
66.9 3 does not influence me.

(53) *Want* to return because I prefer to be in the middle of things rather than on the side-lines.

11.9 1 influences me greatly.
36.2 2 influences me somewhat.
51.7 3 does not influence me.

(54) Which *one* of the statements below best describes your present attitude toward returning to combat duty? (Write only one number in space No. 54.)

12.5 1 would return immediately.
3.0 2 would return after one more month.
3.8 3 would return after two more months.
3.8 4 would return after three more months.
2.7 5 would return after four more months.
2.1 6 would return after five more months.
39.0 7 would return after six more months.
28.0 8 would not return under any circumstances.

Indicate the degree to which each of the following factors influences your feelings *against* returning to combat.

(55) *Do not want* to return because of family ties and obligations.

43.4 1 influences me greatly.
32.0 2 influences me somewhat.
24.6 3 does not influence me.

(56) *Do not want* to return because I feel I have contributed my share in combat.

17.6 1 influences me greatly.
32.0 2 influences me somewhat.
50.4 3 does not influence me.

(57) *Do not want* to return because I feel I could not stand up under the strain of further combat.

14.6 1 influences me greatly.
25.8 2 influences me somewhat.
59.5 3 does not influence me.

(58) To what extent do you feel that your stay in the United States since returning from combat duty has affected your efficiency for future duty?

24.4 1 increased my efficiency considerably.
30.1 2 increased my efficiency somewhat.
28.4 3 no effect on my efficiency.
11.4 4 decreased my efficiency somewhat.
5.1 5 decreased my efficiency considerably.

(59) In your opinion, how long should an officer returned from combat, (and physically qualified for overseas duty) remain in the United States before going back for a second tour of combat duty? (In making your judgment, do not include time in hospital.)

.4 1 less than 1 month.
1.7 2 between 1 and 2 months.
2.5 3 between 2 and 3 months.
1.3 4 between 3 and 4 months.
1.1 5 between 4 and 5 months.
9.3 6 between 5 and 6 months.
10.6 7 between 6 and 7 months.
9.8 8 between 7 and 8 months.
61.7 9 more than 8 months.

(60) How strong is your desire to return to combat duty?

8.9 1 very strong desire to return.
21.4 2 some desire to return.
21.0 3 indifferent as to returning (not for or against).
22.3 4 some desire not to return.
[illegible] 5 very strong desire not to return.

(61) If the AAF makes it possible for you to volunteer for a second tour of combat duty, would you

9.7 1 volunteer for second tour immediately.
3.4 2 volunteer for second tour after one more month.
3.0 3 volunteer for second tour after two more months.
3.2 4 volunteer for second tour after three more months.
1.9 5 volunteer for second tour after four more months.
1.9 6 volunteer for second tour after five more months.
17.2 7 volunteer for second tour after six more months.
58.5 8 decline to volunteer for a second tour.

(62) Some men returned from combat have indicated that certain conditions such as the food, recreation, etc., in the overseas theater were very unsatisfactory. They also indicated that if these conditions were improved, they would have a greater desire to return to combat. Which *one* of the following conditions, if improved, would *increase* your desire to return? Mark only the one number you consider the *most* important. (Write only one number in space No. 62.)

56.1 1 food.
16.1 2 living quarters.
17.8 3 rest and recreational facilities.
3.0 4 sanitary conditions (control of disease).
[illegible] 5 medical care.
3.4 6 longer time between missions.

Irrespective of your past training and experience would you be more willing to return to combat if you were assigned to any one or more of the following ships. Be sure to answer every question.

(63) Single engine fighters.
41.7 1 yes.
57.6 2 no.

(64) Twin engine fighters.
43.0 1 yes.
56.4 2 no.

(65) Attack bombers.
28.2 1 yes.
71.4 2 no.

(66) Medium bombers.
29.2 1 yes.
70.5 2 no.

(67) Heavy bombers.
13.8 1 yes.
85.5 2 no.

(68) Very heavy bombers.
33.5 1 yes.
66.1 2 no.

(69) Reconnaissance.
21.2 1 yes.
78.2 2 no.

(70) Would you be more willing to return overseas if given an assignment not involving frequent combat flying?

57.6 1 yes.
42.0 2 no.

(71) Would you be more willing to return to combat if some of the other members of the crew also had previously completed a tour of combat duty?

61.6 1 yes.
36.6 2 no.

(72) Do you have a permanent grade in the Regular Army, National Guard or Reserve Corps?

6.1 1 yes; an enlisted grade in the Regular Army.
.8 2 yes; a commissioned grade in the Regular Army.
.0 3 yes; an enlisted grade in the National Guard.
.2 4 yes; a commissioned grade in the National Guard.
.2 5 yes; an enlisted grade in the Reserve Corps.
69.3 6 yes; a commissioned grade in the Reserve Corps.
23.5 7 no; do not have any permanent grade.

(73) At the end of the present national emergency, do you desire to:

59.1 1 remain on active duty with the Air Forces (Regular Army)?
37.1 2 return to civilian life in an Army Reserve Officer status?
3.8 3 return to civilian life without retaining Army Reserve Officer status?

(74) If you were given a choice of your future assignment during the present emergency, do you have a specific assignment in mind which you would most prefer?

56.8 1 yes.
33.7 2 no.
9.5 (if your answer is "*yes*," write this specific assignment on the back of the answer sheet after No. 74.)

(75) Have you attended any Army school other than the regular bombardier, navigator and pilot (primary, basic, advanced) schools?

42.6 1 yes.
57.4 2 no.
(If your answer is "yes," write the *type* of school or schools attended on the back of the answer sheet after No. 75).

(76) If it were possible for you to get further training or attend an Army technical school, would you be interested in any one or more of the following. Indicate your choice(s) on the back of the answer sheet after No. 76. List any others you may wish.

1 pilot training.
2 bombardier training.
3 navigator training.
4 engineering school.
5 radio school.
6 weather school.
7 gunnery school.
8 combat intelligence.

(77) If, at some future time, it were possible for the Air Corps to set up a training program which would not necessarily be limited to Army specialties would you be interested in going to school for three (3) to six (6) months?

1 yes.
2 no.
(If your answer is "yes," list in order of preference on back of answer sheet after No. 77 the three (3) courses which you would like to take. Remember that these preferences need not necessarily be related or limited to Army specialties.

Appendix I.2

Gunnery Questionnaire

(Based on 2659 cases unless otherwise indicated)

This questionnaire is to determine and record your experiences and opinions as a returned combat gunner. The Army Air Force is vitally interested in finding out your reaction to your training, to what you have done and seen while in combat, and to your present ideas. This questionnaire is not to be considered a test of any kind but rather as a record of what you have seen, what you have done, and what you think.

You will find that the questions may be answered by indicating the number of one of the alternative answers in the space provided on the answer sheet. Be sure that you write the number indicating your answer in the space corresponding to the question. However, in some of the questions you will find that the alternatives listed do not give your answer. Instructions are given with these questions to write your answer in the space provided on the *back* of your answer sheet. Read the question and all of the alternative answers before answering the question. If none of the alternative answers is satisfactory to you, raise your hand and the administrator will gladly help you. Do not skip any question. Answer each one to the best of your ability.

This booklet of questions will be used again for another returned gunner. Therefore, you are requested not to make any pencil marks upon it or deface it in any way. Write your name, serial number, present post or field and date on the answer sheet in the space provided and then proceed with the first question. There will be no time limit. When you have finished, please raise your hand.

The first question asks "What is your present grade?" Under the question is a list of grades from Master Sergeant to Private. Pick out your present grade and write the number corresponding to it on the list of alternative answers in the space under "1" on your answer sheet. Continue in the same manner for the remainder of the questions.

	Age	Percent		Age	Percent
0		0.1	5	27, 28	11.5
1	20 or below	2.3	6	29, 30	8.1
2	21, 22	18.7	7	31, 32	5.5
3	23, 24	28.3	8	33, 34	2.5
4	25, 26	19.3	9	35 and over	3.3

(Percentage distribution of answers is indicated in margin.)

(1) What is your present grade? (Write the correct number in space No. 1.)

1.5 1 Master Sergeant.
.8 2 First Sergeant.
36.4 3 Technical Sergeant.
56.6 4 Staff Sergeant.
2.7 5 Sergeant.
.7 6 Corporal.
.1 7 Private First Class.
1.1 8 Private.

(2) What is your age in years, to your nearest birthday? (Write it in space No. 2.)

(3) Before you entered the Army, how far along in school did you get?

1.7 1 Less than the eighth grade.
8.1 2 completed the eighth grade.
18.0 3 one or two years in high school or vocational school.
60.3 4 three or four years in high school or vocational school.
5.7 5 one year in college.
3.6 6 two years in college.
1.2 7 three years in college.
1.1 8 four years in college or graduated from college.
.1 9 did post graduate work in college or university.
.0 10 attended or graduated from professional school (dental, medical, etc.)
.2 0

(4) In general, how were your grades in schools you attended before coming into the Army? (Write only one number in the space No. 4.)

0.7 1 superior.
20.4 2 above average.
78.4 3 average.
.5 4 below average.
.1 5 failing.
.04 0

(5) Are you married or single? (Write only one number in Space No. 5.)

54.3 1 single.
.2 2 widowed.
1.8 3 divorced or separated other than by military necessity.
19.3 4 married, *before* going overseas.
.7 5 married, *while* overseas.
23.7 6 married, *since* returning from overseas.

(6) How many living children do you have?

9.6 1 one.
1.2 2 two.
.4 3 three.
.04 4 four.
.0 5 five or more.
87.5 6 none.
1.4 0

(7) What was your reaction at the time you *first* learned that you were assigned to gunnery training? (Write only one number in space No. 7.)

30.2 1 highly pleased.
41.2 2 pleased.
12.8 3 indifferent.
2.4 4 displeased.
.9 5 very unhappy.
9.0 6 never had gunnery training.
3.6 0

(8) In which *one* of the following fields of gunnery do you think you could have benefited most by additional training before being sent into combat? (Write only one number in space No. 8.)

10.5 0
58.3 1 air to air firing.
1.9 2 machine gun ground firing.
9.1 3 turret operation.
.8 4 intercommunication training (intercom).
2.1 5 range estimation.
7.6 6 sights or sighting.
2.9 7 in no fields.
3.0 8 parachute and escape procedures.
3.8 9 other (If not listed above, write what field on the back of the answer sheet after No. 8.)
10 never had gunnery training.

(9) *After* finishing gunnery training, but *before* you were actually assigned to go on a mission, how did you feel about getting into combat?

1.3 0

38.7 1 favored it very much.

21.8 2 favored it moderately.

10.6 3 favored it a little.

14.7 4 indifferent—neither for nor against it.

2.7 5 against it slightly.

.9 6 against it moderately.

.6 7 against it very much.

8.7 8 never had gunnery training.

(10) What was your principal duty during *most* of your time in combat theaters? (Write only one number in space No. 10.)

26.4 1 gunner and mechanical specialist (engineer, crew chief, mechanic, etc.).

21.1 2 gunner and radio or radar specialist.

35.9 3 gunner and armorer.

.9 4 gunner and photographer.

.6 5 gunner and fire control specialist.

.3 6 gunner and remote control turret mechanic.

13.9 7 gunner only.

9.9 0

(11) If you received any training in special fields other than gunnery such as armament, radio, mechanics, etc., either in the United States or overseas, was it sufficient for your needs in combat? (Write only one number in space No. 11.)

21.1 1 sufficient in all phases.

43.8 2 sufficient in most phases.

13.1 3 sufficient in a few phases.

5.7 4 entirely insufficient.

15.5 5 did not receive training in fields other than gunnery.

.7 0

(12) How much use did you make *in combat* of the training you received (both in the United States and overseas) in the special field other than gunnery such as armament, radio operation, airplane mechanics, and other similar fields. (Write only one number in space No. 12.)

1.4 0

14.3 1 very extensive use.

28.0 2 wide use.

42.0 3 some use.

7.0 4 no use.

7.3 5 received no training other than gunnery.

(13) After you were assigned to a combat crew, which *one* of the following reasons was the strongest in keeping you there? (Write only one number in space No. 13.)

13.3 1 thought it was the best way to strike an active blow against the enemy.

46.4 2 desired to do a good job and contribute my share.

4.1 3 didn't want to let the other members of the crew down by quitting.

9.3 4 liked the excitement of combat gunnery.

2.1 5 performed the duties because I didn't want to be called a quitter.

21.3 6 preferred flying duty to ground duty.

1.6 7 wanted the extra money from flight duty.

1.1 8 performed the duties only because I was ordered to do so.

.5 0

(14) Judging from your experience in combat from which one of the following age periods do the *most* successful combat gunners come? Make your judgement on the basis of the performance of the gunner in combat. (Write only one number in space No. 14.)

4.7	1	18, 19.
27.7	2	20, 21.
40.1	3	22, 23.
19.7	4	24, 25.
3.9	5	26, 27.
1.4	6	28, 29.
1.1	7	30 or over.
1.5	0	

(15) How long were you overseas.

.4	1	2 months or less.
.9	2	3 or 4 months.
5.7	3	5 or 6 months.
16.7	4	7 or 8 months.
34.2	5	9, 10, or 11 months.
19.8	6	12, 13, or 14 months.
7.6	7	15, 16, or 17 months.
4.9	8	18, 19, or 20 months.
9.0	9	21 or more months.

(16) How much time was there between your *first* combat mission and the completion of your *last* mission?

4.2	1	2 months or less.
9.3	2	3 or 4 months.
23.7	3	5 or 6 months.
24.4	4	7 or 8 months.
23.8	5	9, 10, or 11 months.
7.9	6	12, 13, or 14 months.
2.6	7	15, 16, or 17 months.
1.5	8	18, 19, or 20 months.
1.5	9	21 or more months.
1.1	0	

(17) How many combat missions did you complete?

6.1	1	10 or less.
3.9	2	11 to 20.
23.6	3	21 to 25.
11.5	4	26 to 30.
10.5	5	31 to 40.
27.5	6	41 to 50.
12.2	7	51 to 60.
1.6	8	61 to 70.
1.8	9	More than 70.
1.5	0	

(18) Did you fly more or less than the regularly required number of missions or hours for your theater and type of plane? (Write only one number in space No. 18.)

10.8	1	flew less than the required number of missions or hours.
51.4	2	flew exactly the number of missions or hours.
16.9	3	flew more than the required number of missions or hours.
20.3	4	there was no exact policy on the number of missions or hours required or do not know.
5.5	0	

(19) In what theater did you fly all or *most* of your combat missions?

32.3 1 European (England).
35.7 2 Mediterranean (Italy, North Africa, Middle East).
18.2 3 Pacific (Central, South, Southwest).
3.2 4 China-Burma-India.
2.5 5 Caribbean.
3.9 6 Alaska.
.9 7 North Atlantic—Greenland.
.3 8 other (If you flew in a theater not listed here, write its name on the back of the answer sheet after No. 19.)

(20) In what one type of plane did you fly all or most of your combat missions? (Write only one number in space No. 20.)

55.5 1 B-17.
21.6 2 B-24.
13.0 3 B-25.
6.4 4 B-26.
2.6 5 A-20.
.1 6 B-34.

(21) As an aerial gunner, what was your position in the plane *most* of the time? (Write only one number in space No. 21.)

3.2 1 nose gun or nose turret.
26.5 2 top gun or top turret.
10.3 3 radio operator's gun.
13.7 4 bottom gun or ball turret.
24.4 5 waist gun.
20.4 6 tail gun or tail turret.
.9 7 other (If you fired from a position not listed here, write location on the back of the answer sheet after No. 21.)
.6

(22) From your experiences in combat, which gunner's position would you say is the most dangerous? (Write only one number in space No. 22.)

9.6 1 nose gun or nose turret.
4.7 2 top gun or top turret.
.8 3 operator's gun.
8.2 4 bottom gunner ball turret.
18.5 6 waist gun.
14.3 6 tail gun.
.2 7 other (If it is a position not listed above, write its location on the line after No. 22.)
43.2 8 all gun positions are equally dangerous.
.6 0

(23) Were you ever in a crash landing, forced landing, or "ditching" that caused injury to any person? (Write only one number in space No. 23.)

.07 0
23.7 1 yes; in friendly territory.
4.3 2 yes; in enemy territory, or territory not controlled by friendly groups.
1.7 3 yes; in both friendly and enemy territory at different times.
56.7 4 no; never had a crash landing.
13.0 5 had a crash landing, forced landing or ditching but no one injured.

(24) On what proportion of your missions did you meet enemy opposition of some kind? Opposition includes flak, fighters, etc. (Write only one number in space No. 24.)

0.3 0
3.0 1 never.
1.8 2 rarely.
2.1 3 less than half of the missions.
11.0 4 half or more of the missions.
46.5 5 almost every mission.
35.2 6 every mission.

(25) Did you ever "bail-out" on a com' -t fight? (Write only one number in space No. 25.)

3.3 1 yes; over friendly territory.
4.3 2 yes; over enemy territory, or territory not controlled by friendly troops (by authority of the Intelligence office, returnee may give this answer if it is true).
.1 3 yes; over both friendly and enemy territory at different times.
92.2 4 no; never bailed out.
.2 0

(26) Were you ever wounded in combat? (Write only one number in space No. 26, if you were wounded more than once, answer for the worst one.

0.3 0
11.9 1 slight wound; not in hospital, or in hospital less than one week.
3.2 2 moderately severe wound; in hospital up to 6 weeks.
2.6 3 severe wound; in hospital more than 6 weeks.
80.2 4 was never wounded.
1.8 5 received mo 'erately severe wound but was not hospitalized.

(27) While on foreign service, were you ever *injured* in connection with aircraft while *not* in combat? (Write only one number in space No. 27. If you were injured more than once, answer for the worst injury.)

10.1 1 slight injury; not in hospital, or in hospital less than 1 week.
1.2 2 moderately severe injury; in hospital up to 6 weeks.
.8 3 severe injury; in hospital more than 6 weeks.
87.7 4 was never injured.
.2 0

(28) How much were you bothered by air-sickness while in combat?

76.0 1 never.
20.2 2 on only a few flights.
2.9 3 frequently.
.6 4 on nearly every flight.
.3 5 on every flight.
.2 0

(29) While on foreign service, were you ever— (Mark the number of the final statement that applies to you. Write only one number in space No. 29.)

2.7 1 sent to a hospital because of operational fatigue or exhaustion.
22.5 2 sent to a rest camp because of operational fatigue.
12.1 3 given furlough(s) for reason of operational fatigue.
7.8 4 diagnosed as having flying fatigue or opera 'onal fatigue, but not treated for it.
21.3 5 sent to a rest camp routinely, not for operational fatigue.
33.0 6 never had operational fatigue nor went to . rest camp.
.5 0

(30) Were any of your crew mates killed in the planes in which you flew? (Write only one number in space No. 30.)

14.4 1 1 was killed.
3.5 2 2 were killed.
2.7 3 3 were killed.
1.1 4 4 were killed.
.6 5 5 were killed.
.7 6 6 were killed.
.5 7 7 were killed.
.9 8 8 were killed.
76.3 9 none was killed.
.2 0

(31) Were any of your crew mates wounded in the planes in which you flew? (Write only one number in space No. 31.)

35.4 1 1–2 were wounded.
15.6 2 3–4 were wounded.
3.8 3 5–6 were wounded.
1.7 4 7–8 were wounded.
.8 5 9–10 were wounded.
.6 6 11–12 were wounded.
.2 7 13–14 were wounded.
.1 8 15–16 were wounded or more.
41.5 9 none were wounded.

(32) How did the death or wounding of your crew mates affect your desire to remain on combat duty?

6.2 0
5.9 1 greatly increased my desire for combat duty.
12.4 2 somewhat increased my desire for combat duty.
35.1 3 didn't affect my desire for combat duty.
14.6 4 somewhat decreased my desire for combat duty.
5.0 5 greatly decreased my desire for combat duty.
20.7 6 none of my crew mates were killed or wounded.

(33) How many of your close friends, *not crew mates,* were missing, killed or wounded? (Those who flew in the same plane are not to be considered. Write only one number in space No. 33.)

35.9 1 half or more of them were missing, killed or wounded.
23.2 2 between a quarter and a half were missing, killed or wounded.
22.0 3 several were missing, killed or wounded, but less than a quarter.
14.9 4 only a few were missing, killed or wounded.
3.6 5 no close friend was missing, killed or wounded.
.4 0

(34) In general, how well did you like your job as an aerial gunner while *in combat?*

32.2 1 liked it very much.
33.5 2 liked it somewhat.
22.9 3 indifferent—neither liked or disliked it.
7.6 4 disliked it somewhat.
2.8 5 disliked it very much.
.9 0

(35) What was the *principal* reason for your being returned to the United States from an overseas theater? (Write only one number in space No. 35.)

63.2 1 rotation policy; having flown the required number of missions or hours.
19.9 2 returned because of flying fatigue, after completing about the usual number of missions in my theater.

1.3 3 returned because of flying fatigue or operational fatigue without completing the usual number of missions.
2.6 4 wounded in action or injured in connection with aircraft.
1.1 5 illness, injury, or operation not connected with aircraft.
4.9 6 escaped prisoner of war or evader of capture (On instructions of Intelligence Officer, returnees may give this reason if it is true.)
6.1 7 other reason. (Write what on the back of the answer sheet after No. 35.)
0.9 0

(36) How long have you been in this country since you returned from duty overseas?

0.4 1 less than 1 month.
3.7 2 1 month.
6.4 3 2 months.
9.6 4 3 months.
10.0 5 4 months.
12.5 6 5 months
13.0 7 6 months.
8.0 8 7 months
36.3 9 more than 7 months.

(37) How do you feel about the duration of the furlough and rest period given you upon your return to the United States and before you were reassigned to active duty?

43.5 1 much too short.
31.0 2 just a little too short.
22.1 3 about right.
.6 4 just a little too long.
.1 5 much too long.
2.3 6 no furlough or rest period given me.
.4 0

(38) What is your present MOS? (Military Occupation Specialty.)

2.1 1 do not know.
8.6 2 611-aerial gunner.
31.8 3 612-airplane armorer.
21.5 4 748-army airplane mechanic-gunner.
15.3 5 757-AAF radio operator-mechanic-gunner.
.1 6 580-remote control turret mechanic.
.0 7 960-remote control turret mechanic.
.0 8 940-army aerial photographer.
20.2 9 other. (If not listed above write MOS number and job title on the back of the answer sheet after no. 38.)
.4 0

(39) *Since* you returned to the United States how many jobs have you had on which you have worked a week or longer?

59.1 1 1 job.
21.8 2 2 jobs.
7.0 3 3 jobs.
2.1 4 4 jobs.
.8 5 5 jobs.
.1 6 6 jobs.
.1 7 7 jobs.
.4 8 8 jobs or more.
8.1 9 no job.
.2 0

(40) Which *one* of the following *best* describes your *major* duty assignment since you have returned to the United States?

57.1 1 instructor on flying status.
20.7 2 ground instructor in gunner, armament, radio, etc.
2.5 3 instructor other than 1 or 2.
9.7 4 armament, radio, mechanic, etc. (ground duty but not instructing.)
.7 5 clerical or administrative.
.1 6 public relations.
.5 7 in hospital.
4.4 8 no assignment.
3.8 9 other. (Write what, on back of answer sheet after No. 40.)
.5 0

(41) In your opinion, how well is the Army now making use of your ability and experience other than in combat gunnery, military or pre-war experience such as armament, radio, clerical, administrative, etc.

25.1 1 best use of my ability and experience.
46.7 2 good use of my ability and experience.
16.1 3 very little use of my ability and experience.
7.3 4 no use of my ability and experience.
4.5 5 no assignment has been given to me.
.4 0

(42) Which one of the following phrases *best* describes your satisfaction with your present assignment?

22.1 1 very satisfactory.
45.7 2 satisfactory.
14.3 3 indifferent.
10.0 4 unsatisfactory.
6.8 5 very unsatisfactory.
.9 0

(43) How well do you think you are performing the duties of your present assignment?

5.2 1 outstandingly.
65.5 2 well.
25.3 3 acceptably.
1.5 4 poorly.
1.6 5 very poorly.
1.8 0

(44) Do you feel that your services as a *combat* gunner are of more value in winning the war than the services you are performing in your *present* job?

12.8 1 services as a combat gunner very much more valuable.
10.1 2 services as a combat gunner somewhat more valuable.
34.7 3 services as a combat gunner of equal value to services on present job.
19.4 4 services as a combat gunner somewhat less valuable.
20.0 5 services as a combat gunner very much less valuable.
2.9 6 0

(45) Which one of the following conditions in the United States are you *most dissatisfied* with? (Answer only one number in space No. 45.)

1.0 0
18.6 1 the attitude of labor in this country.
5.9 2 the attitude of business and industrial management in this country.
11.7 3 the general attitude of civilians to military personnel.
15.0 4 the nature of military courtesy, discipline and training in this country.
2.1 5 rationing.

31.1 6 the attitude of military personnel who have never been overseas.
1.9 7 my own family relationships.
12.5 8 not dissatisfied with any condition in the United States.

(46) How strong is your desire to remain on flying status?

22.8 1 very strong.
24.9 2 strong.
22.4 3 indifferent.
6.3 4 somewhat opposed to it.
4.3 5 strongly opposed to it.
18.9 6 not on flying status at present time.
.3 0

(47) Which *one* of the following statements *best* describes your feelings relative to returning to combat duty as a gunner?

.9 1 want to return to square accounts because of my friends who have been wounded and killed overseas.
6.3 2 want to return because I prefer duty as a combat gunner to the duty to which I am now assigned.
5.8 3 want to return because I am dissatisfied with conditions as I have found them here in the United States.
.2 4 want to return in order to obtain increased pay.
12.6 5 do not care whether or not I return to combat.
20.9 6 do not want to return because of family ties and obligations.
16.3 7 do not want to return because I feel I have contributed my share in combat.
36.2 8 do not want to return because I feel I could not stand up under the strain of further combat.
.7 0

(48) Which *one* of the statements below best describes your *present attitude* toward returning to combat duty as a gunner? (Write only one number in space No. 48.)

4.1 1 eager to return immediately.
12.0 2 would return at any time I am needed.
.8 3 would return if I could first have advanced training in gunnery.
.7 4 would return after I have been in this country about 3 months.
1.6 5 would return after I have been in this country about 6 months.
9.2 6 would want to return only if the war continues longer than is generally expected.
48.8 7 would want to return only after all available trained gunners in the United States have been sent to combat.
22.0 8 would not want to return under any circumstances.
.6 0

(49) Some men returned from combat have indicated that certain conditions such as the food, recreation, etc., in the overseas theater were very unsatisfactory. They also indicated that if these conditions were improved they would have a greater desire to return to combat. Which *one* of the following conditions, if improved, would *increase* your desire to return? Mark only one which you consider *most* important. (Write only one number in space No. 49.)

48.5 1 food.
13.1 2 living quarters.
21.2 3 rest and recreational facilities.
4.8 4 sanitary conditions (control of disease.)
.7 5 medical care.
6.0 6 longer time between missions.
5.6 0

50) Are you *now* more desirous of returning for a second tour of combat duty than you were at the time you *first arrived* back in the United States.

6.3 1 have much more desire now than when I first returned to the United States.

10.0 2 have more desire now than when I first returned to the United States.

26.8 3 have about the same desire now as when I first returned to the United States.

23.6 4 have less desire now than when I first returned to the United States.

32.6 5 have much less desire now than when I first returned to the United States.

.7 0

(51) To what extent do you feel that your stay in the United States since returning from combat duty has affected your efficiency for future combat duty?

14.5 1 undoubtedly increased my efficiency.

23.2 2 probably increased my efficiency.

35.0 3 no effect.

17.3 4 probably decreased my efficiency.

9.6 5 undoubtedly decreased by efficiency.

.4 0

(52) In your opinion, how do most returned combat gunners feel about returning for another tour as combat gunners?

.9 1 very willi'g to return.

5.2 2 willing to return.

10.9 3 indifferent as to returning.

53.8 4 not anxious to return.

28.8 5 definitely do not want to return.

.3 0

(53) How strong is your desire to return to combat as a gunner?

5.0 1 very strong desire to return.

9.8 2 some desire to return.

15.7 3 indifferent as to returning (not for or against)

27.5 4 some desire not to return.

40.1 5 very strong desire not to return.

1.8 0

(54) Do you expect to volunteer for a second combat tour of duty as a gunner?

4.7 1 plan to volunteer for second tour immediately.

1.8 2 plan to volunteer for second tour after 1 more month.

1.5 3 plan to volunteer for second tour after 2 more months.

1.3 4 plan to volunteer for second tour after 3 more months.

1.3 5 plan to volunteer for second tour after 4 more months.

.7 6 plan to volunteer for a second tour after 5 more months.

4.3 7 plan to volunteer for second tour after 6 more months.

81.4 8 do not plan to volunteer for a second tour.

2.9 0

(55) In your opinion, how long should a gunner returned from combat remain in the United States before going back for a second tour of combat duty? (In making your judgment, do not include time in hospital.)

1.5 1 less than 1 month.

1.4 2 between 1 and 2 months.

2.3 3 between 2 and 4 months.

1.5 4 between 3 and 4 months.

5.5	5	between 5 and 6 months.
10.9	6	between 6 and 7 months.
4.4	7	between 7 and 8 months.
7.5	8	between 9 and 10 months.
60.2	9	more than 10 months.
3.8	0	

If you were to be returned to combat as a gunner, do you feel that you would need more training in specific phases of gunnery before being sent overseas? (If you do not feel in need of additional training, write "1" in all three spaces, 56, 57 and 58.)

41.6	1	do not feel in need of additional training.
27.3	2	would like more training in air-to-air firing.
8.1	3	would like more training in turret operation.
2.3	4	would like more training in range estimation.
6.7	5	would like more training in sights and sighting.
.4	6	would like more training in intercommunication training (intercom.).
8.5	7	would like more training in other air crew duties, crewmanship, etc.
2.8	8	would like more training in Jam Handy, Waller, Trainer, etc.
2.3	0	

(56) Record in space No. 56 the number of your first choice (the type of training you need most).

(57) Record in space No. 57 the number of your second choice or requirement.

(58) Record in space No. 58 the number of your third choice or requirement.

Some gunners received training in a specialty other than gunnery such as armament, radio operation, etc. Indicate in the next two questions whether you feel that your present ability in the specialty other than gunnery which you may have is sufficient for meeting the needs of combat operations, or whether you think a refresher course would be required.

(59) In which one of the specialties listed below do you think you need a refresher course before going back to combat? (Write only one number in space No. 59.)

14.8	1	received no special training other than gunnery.
36.7	2	do not need a refresher course in my specialty.
9.6	3	need a refresher course in radio operation.
10.4	4	need a refresher course in mechanics.
16.8	5	need a refresher course in armament.
6.5	6	need a refresher course in airplane mechanics.
1.0	7	need a refresher course in photography.
1.5	8	other (if not listed above, write in specialty on back of answer sheet after No. 59).
2.8	0	

(60) In which one of the specialties listed below do you consider yourself qualified for meeting the needs of combat?

15.7	1	received no special training other than gunnery.
19.8	2	need a refresher course in my specialty.
14.0	3	feel qualified for combat in radio operation.
11.0	4	feel qualified for combat in mechanics.
27.4	5	feel qualified for combat in armament.
7.1	6	feel qualified for combat in airplane mechanics.
.4	7	feel qualified for combat in photography.
1.3	8	other (if not listed above write in specialty on back of answer sheet after No. 60).
3.2	0	

(61) Since returning from combat, do you become easily exhausted or all tired out?

48.9 1 often.
43.6 2 sometimes.
7.3 3 never.
.2 0

(62) Since returning from overseas duty, have you felt sick to your stomach or felt that you had to vomit?

12.4 1 often.
45.8 2 sometimes.
41.6 3 never.

(63) Do you now find it difficult to concentrate on tasks that didn't bother you before you went to combat?

25.5 1 often.
50.8 2 sometimes.
23.7 3 never.

(64) Do you have shaking or trembling hands or knees or muscular twitches?

24.4 1 often.
51.3 2 sometimes.
24.1 3 never.
.1 0

(65) Are you now more easily made grouchy or more easily irritated than you were before you went to combat?

58.9 1 more.
31.8 2 just as often.
9.1 3 less.
.1 0

(66) Do thoughts of your combat experience produce feelings of fear?

14.3 1 often.
53.0 2 sometimes.
32.6 3 never.
.1 0

(67) Since returning from overseas duty are you easily confused or "rattled"?

16.2 1 often.
55.6 2 sometimes.
28.1 3 never.
.1 0

(68) Do you worry about things or conditions that you probably can't change anyway?

23.4 1 often.
53.8 2 sometimes.
22.7 3 never.

(69) Do you have fears which you cannot seem to drive out of your mind?

13.7 1 often.
41.9 2 sometimes.
44.2 3 never.
.1 0

Distribution of Health Scores

Sum of answer to questions 61 to 76 inclusive

	Score	Percent
1	22 and below	7.8
2	23, 24, 25	10.6
3	26, 27, 28	14.9
4	29, 30, 31	16.3
5	32, 33, 34	15.1
6	35, 36, 37	11.8
7	38, 39, 40	10.4
[illegible]	41, 42, 43	7.8
9	44 and above	

(70) Are you now bothered with sleepless nights?

19.3 1 often.
52.3 2 sometimes.
28.3 3 never.

(71) Are you now restless or not able to sit still?

42.4 1 often.
45.9 2 sometimes.
11.8 3 never.

(72) Do loud or sudden sounds make you jumpy?

40.2 1 often.
46.1 2 sometimes.
13.8 3 never.

(73) Do you feel "blue" or depressed?

21.2 1 often.
63.6 2 sometimes.
15.2 3 never.

(74) Do you become hesitant and so uncertain of yourself that you are unable to make a decision as quickly as you think you should?

15.3 1 often.
53.9 2 sometimes.
30.8 3 never.

(75) Do you relive the "close calls" which you experienced in combat? (Not tallied—column 75 used to code health score.)

1 often.
2 sometimes.
3 never.
4 did not have "close calls."

(76) Since returning from combat do you have nightmares?

10.0 1 often.
45.3 2 sometimes.
43.9 3 never.
.7 0

703330—47—24

(77) In your opinion, how long does it require to *rest up* after completing a tour of duty as a gunner in a combat crew? In making your judgment do not consider cases requiring an extended stay in a hospital. (Write only one number in space No. 77).

1.0 1 less than 2 weeks.
1.8 2 2 or 3 weeks.
4.4 3 1 month.
5.2 4 2 months.
6.3 5 3 months.
3.4 6 4 months.
1.6 7 5 months.
18.9 8 6 months.
54.1 9 more than 6 months.
3.2 0

(78) If you were given a choice today between assignment to a second tour of duty as a gunner or assignment as a basic soldier, which assignment would you prefer?

39.9 1 assignment to a second tour as a gunner.
55.4 2 assignment as a basic soldier.
4.6 0

(79) After you returned from overseas duty, were you processed in a Redistribution Center between 24 May 1944 and 6 Aug. 1944? The three Redistribution Centers are at Atlantic City, N. J., Miami Beach, Fla., and Santa Monica, Calif.

25.4 1 not processed in Redistribution Center.
11.4 2 processed at Atlantic City between 24 May 1944 and 6 August 1944.
13.1 3 processed at Miami Beach between 24 May 1944 and 6 August 1944.
6.1 4 processed at Santa Monica between 24 May 1944 and 6 August 1944.
42.6 5 processed at Atlantic City, Miami Beach or Santa Monica but not during period 24 May 1944 to 6 August 1944.
1.4 0

(80) Following is a list of courses offered at most gunnery schools. Write down the number of each of the courses in which you have not received instruction. Be sure you indicate every course which you have not taken. (Write the numbers on the back of the answer sheet after No. 80).

28.7 1 position firing.
52.8 2 gun camera missions.
19.6 3 preventive maintenance.
21.3 4 basic deflection range.
30.7 5 shotgun turrets.
46.0 6 Jam Handy training.
8.1 7 turret operation.
14.6 8 training on the job and not in gunnery school.
18.9 9 received instructions in all courses listed.
2.5 10 never had gunnery training.
3.8 0

(81) If it became necessary for you to make one of the following choices which one would you select?

1. I no longer want to remain on flying status but prefer reassignment to ground duty in the Air Forces in this country.
1. I wish to remain on flying status and am willing to return to combat for a second tour after I have received further training even though during part of this training I do not receive flight pay.
3. I wish to remain on flying status and want immediate assignment to a second tour of combat duty.

(82) State briefly what promises were made to you when you came through the Redistribution Center, concerning assignment, type of work, possible location, etc. Tell to what extent these promises were met or fulfilled.

(83) Would you be more willing to return to combat if you were assigned to a B-29?

1. Yes.
2. No.

(84) If, at some future time, it were possible for the Air Corps to set up a training program which would not necessarily be limited to army specialties would you be interested in going to school for three (3) to six (6) months?

1. Yes.
2. No.

If your answer is yes, list in order of preference on back of answer sheet the three courses which you would like to take. Remember that these preferences need not necessarily be related or limited to army specialties.

(85) At present, how much are you bothered by air sickness?

1. never.
2. only on a few flights.
3. frequently.
4. on nearly every flight.
5. on every flight.
6. not on flying status.

Subject Index

★ U. S. GOVERNMENT PRINTING OFFICE: 1946—703330